

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA



**DISSERTAÇÃO DE MESTRADO**

*Comparação via Simulação dos Estimadores Clássicos e  
Bayesianos no Modelo de Coeficientes Aleatórios  
para Dados Longitudinais*

Vanessa Bielefeldt Leotti

Orientador: Prof. Dr. Álvaro Vigo

Porto Alegre, dezembro de 2007.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA



**DISSERTAÇÃO DE MESTRADO**

*Comparação via Simulação dos Estimadores Clássicos e  
Bayesianos no Modelo de Coeficientes Aleatórios  
para Dados Longitudinais*

Vanessa Bielefeldt Leotti

**Orientador: Prof. Dr. Álvaro Vigo**

A apresentação desta dissertação é exigência do Programa de Pós-graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Mestre.

Porto Alegre, Brasil.  
2007

## **BANCA EXAMINADORA**

Prof. Dra. Patrícia Klarmann Ziegelmann, Departamento de Estatística, UFRGS.

Prof. Dra. Suzi Comey, Programa de Pós-Graduação em Epidemiologia, UFRGS.

Prof. Dr. João Riboldi, Programa de Pós-Graduação em Epidemiologia, UFRGS.

## AGRADECIMENTOS

A Deus, por sempre se manifestar na minha vida quando mais preciso.

À minha família, pelo incentivo, carinho, compreensão e paciência, e  
ao meu noivo, meu maior companheiro e incentivador,  
sem vocês nada disso seria possível.

Ao prof. Álvaro Vigo, pelo incentivo durante o mestrado, pelo apoio  
fundamental durante o concurso e pelo que está fazendo agora, na reta final.

À prof. Patrícia Ziegelmann, pelo apoio durante a graduação e o mestrado.  
E por último, a todos os amigos, que de uma forma ou de outra contribuíram  
para mais essa conquista de minha vida.

## **ABREVIATURAS E SIGLAS**

AIC - Akaike's Information Criterion

ANOVA - Analysis of Variance

AR(1) - First-Order Auto-regressive

BIC - Bayesian Information Criterion

BLUE - Best Linear Unbiased Estimator

BLUP - Best Linear Unbiased Predictor

BUGS - Bayesian inference Using Gibbs Sampling

CCI - Coeficiente de Correlação Intraclasse

CODA - Convergence Diagnostics and Output Analysis

DIC - Deviance Information Criterion

EBLUE - Empirical Best Linear Unbiased Estimator

EBLUP - Empirical Best Linear Unbiased Predictor

EQM - Erro Quadrático Médio

GLM - Modelo Linear Geral

HDR - Highest Density Region

IGLS - Iterative Generalized Least Squares

MCMC - Markov Chain Monte Carlo

MINQUE - Minimum Norm Quadratic Unbiased Estimator

MIVQUE - Minimum Variance Quadratic Unbiased Estimator

ML - Maximum Likelihood

REML - Restricted Maximum likelihood

RIGLS - Restricted Iterative Generalized Least Squares

UN - Unstructured

VC - Variance Components

## RESUMO

Frequentemente em pesquisas médicas ou epidemiológicas, múltiplas medidas de um mesmo sujeito são tomadas ao longo do tempo, caracterizando um estudo com dados longitudinais. Nos últimos anos, a técnica estatística que tem sido mais utilizada para a análise desses estudos é o modelo misto, pois este permite incorporar a provável correlação das observações de um mesmo indivíduo e é flexível para lidar com situações de desbalanceamento e dados faltantes. Um modelo misto muito utilizado nesses casos é o modelo de coeficientes aleatórios, que permite que a relação da variável resposta com o tempo seja descrita através de uma função matemática. Há duas abordagens para se analisar o modelo misto: clássica e Bayesiana. Para se aplicar o método Bayesiano nesse caso, deve-se utilizar um dos métodos MCMC, já que a distribuição a posteriori não é analiticamente derivável. O método MCMC utilizado neste trabalho foi o Amostrador de Gibbs. As duas abordagens podem levar a resultados diferentes, deixando o pesquisador em dúvida sobre qual método utilizar. Poucos estudos compararam a abordagem clássica e a Bayesiana para a análise de dados longitudinais via modelos mistos. O objetivo deste trabalho é proceder a um estudo de simulação, para comparar estas abordagens em termos de vício e precisão. Compararam-se os estimadores Bayesianos média, moda (estimada não-parametricamente) e mediana a posteriori e o estimador clássico obtido pelo método REML. Simulou-se um modelo que envolve efeito de tempo e de uma covariável denominada tratamento, assumindo a estrutura Componentes de Variância para as duas matrizes de covariâncias do modelo misto. Diferentes configurações de tamanho de amostra e desbalanceamento foram adotadas, para avaliar o desempenho dos métodos frente a essas situações. Em cada uma das configurações, foram realizadas 1000 replicações no software R. As maiores diferenças encontradas foram em relação a alguns componentes de variância, sendo que pelo menos um dos estimadores Bayesianos apresentou erro quadrático médio menor do que o estimador clássico em todas as configurações. Não foi possível identificar um único estimador Bayesiano como sendo o melhor para todos os casos estudados. A moda mostrou-se um estimador com boas propriedades em algumas situações e por isso sugere-se que o mesmo seja implementado no software WinBUGS.

## ABSTRACT

Frequently in medical or epidemiologic research, multiple measures of the same subject are observed over the time. This characterizes a study with longitudinal data. In the last years, the statistical technique that has been used in the analysis of these studies is the mixed model, because it permits to model the probable correlation between the observations of the same individual, and it handles well with situations of unbalanceament and missing data. A mixed model that has been used in these cases is the random coefficients model. It allows the relationship between the response variable and the time to be described through a mathematical function. There are two approaches to analyze mixed models: the classical and the Bayesian methods. In order to apply the Bayesian inference in this case, the MCMC methods have to be used, because the posterior distribution is not analytically derivable. The MCMC method used in this work was the Gibbs Sampler. The two approaches may produce different results, and because of that the researcher may be in doubt of what method to use. In the literature, there are few studies that compare the classical and the Bayesian approach in the analysis of longitudinal data using mixed models with an analytical way. So, the objective of this work is to proceed a simulation study to compare the Bayesian and classical inferences in terms of bias and precision. The Bayesian estimators posterior mean, (non-parametric) mode and median, and the classical estimator obtained by the REML method were compared. We simulated a model that involves the effects of time and a covariable named "treatment", and that assumes a Variance Components structure for the two covariance matrices of the mixed model. Different configurations of sample size and unbalanced data were adopted, to see the performance of the methods with these situations. In each of these configurations, 1000 replications were proceeded in the R software. The main differences founded were in the estimation of the variance components of the random effects. For these parameters, one of the Bayesian estimators had lower mean square error than the classical in all of the configurations. It was not possible to identify one best Bayesian estimator in all the cases studied. The mode seems an estimator with good properties in some situations and because of that it should be implemented in the WinBUGS software.

## APRESENTAÇÃO

Este trabalho consiste na dissertação de mestrado intitulada **Comparação via Simulação dos Estimadores Clássicos e Bayesianos no Modelo de Coeficientes Aleatórios para Dados Longitudinais**, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 20 de dezembro de 2007. O trabalho é apresentado em três partes, na ordem que segue:

- Introdução, Revisão da Literatura e Objetivos;
- Artigo;
- Conclusões e Considerações Finais.

Documentos de apoio, incluindo o Projeto de Pesquisa, códigos R para simulação, exemplo de gráfico de convergência e de posteriori estimada e descrição do software WinBUGS, estão apresentados nos anexos.

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>10</b>
<b>2 REVISÃO DA LITERATURA</b> .....	<b>13</b>
2.1 MÉTODOS MCMC .....	15
2.1.1 <i>Cadeias de Markov</i> .....	17
2.1.2 <i>Algoritmo de Metropolis-Hastings</i> .....	18
2.1.3 <i>Amostrador de Gibbs</i> .....	20
2.1.3.1 Definição e propriedades.....	20
2.1.3.2 Geração da amostra, valores iniciais e convergência.....	22
2.1.3.3 Análise dos resultados da simulação.....	23
2.2 ABORDAGEM CLÁSSICA PARA MODELOS MISTOS NORMAIS .....	25
2.2.1 <i>Formulação do modelo misto</i> .....	26
2.2.2 <i>Estimação</i> .....	28
2.2.2.1 Estimação dos efeitos fixos e aleatórios.....	29
2.2.2.2 Estimação dos componentes de variâncias .....	32
2.2.3 <i>Modelos para dados longitudinais</i> .....	34
2.2.3.1 Modelos de padrão de covariância.....	36
2.2.3.2 Modelo de coeficientes aleatórios.....	37
2.3 ABORDAGEM BAYESIANA PARA MODELOS MISTOS NORMAIS .....	39
2.3.1 <i>Conceitos básicos de inferência Bayesiana</i> .....	40
2.3.2 <i>Análise Bayesiana para modelos mistos normais</i> .....	44
2.3.3 <i>Distribuições condicionais completas do modelo de coeficientes aleatórios</i> .....	47
<b>3 REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>52</b>
<b>4 OBJETIVOS</b> .....	<b>54</b>
<b>5 ARTIGO</b> .....	<b>55</b>
<b>6 CONCLUSÕES E CONSIDERAÇÕES FINAIS</b> .....	<b>74</b>
<b>7 ANEXOS</b> .....	<b>76</b>
7.1 ANEXO 1 – PROJETO DE PESQUISA.....	77
7.2 ANEXO 2 – CÓDIGOS R PARA A SIMULAÇÃO.....	84
7.3 ANEXO 3 – EXEMPLO DE GRÁFICO DE CONVERGÊNCIA E DE POSTERIORI ESTIMADA POR KERNEL .....	97
7.4 ANEXO 4 – DESCRIÇÃO DO WINBUGS.....	99

## 1 INTRODUÇÃO

Na pesquisa médica e epidemiológica, é cada vez mais comum se encontrar estudos que utilizam dados longitudinais. Esse tipo de dados ocorre quando múltiplas medidas de um desfecho são tomadas em uma mesma unidade experimental (em geral, indivíduos) em várias ocasiões no tempo. Esses estudos permitem ao pesquisador descrever o comportamento da variável resposta ao longo do tempo, verificar se existe tendência de crescimento ou decrescimento, comparar a eficácia de diferentes intervenções ao longo do tempo, bem como comparar diferenças entre grupos.

Nos últimos anos, a técnica estatística que tem sido mais utilizada para a análise de dados longitudinais é o modelo misto. Esse modelo compreende efeitos fixos e aleatórios em sua formulação e assume distribuição multivariada tanto para seus efeitos aleatórios como para seus erros. Esta última característica permite que dados que não são independentes sejam analisados. Outra propriedade importante do modelo misto é que ele é flexível para analisar situações com dados faltantes e desbalanceamento. Em dados longitudinais, três tipos de desbalanceamento podem ocorrer: número diferente de indivíduos em cada grupo de comparação, número diferente de medidas repetidas tomadas em cada indivíduo e observações realizadas em ocasiões no tempo diferentes.

Há dois casos especiais do modelo misto que são úteis para se analisar dados longitudinais. O primeiro é o modelo de padrão de covariância, onde se assume um padrão para a matriz de variâncias e covariâncias do erro aleatório. Este modelo geralmente é utilizado quando as medidas repetidas são tomadas em intervalos de tempo fixados. O segundo é o modelo de coeficientes aleatórios, onde se ajusta uma função matemática para descrever o comportamento da variável resposta ao longo do tempo. Este modelo pode ser aplicado facilmente quando os intervalos de tempo entre as medidas não são fixos, e pode-se introduzir a correlação das medidas de um mesmo indivíduo permitindo que os efeitos aleatórios tenham uma distribuição multivariada.

A classe dos modelos mistos pode ser analisada por duas abordagens estatísticas: a clássica e a Bayesiana. De uma maneira geral, a principal diferença entre inferência Bayesiana e inferência clássica é que a primeira considera os parâmetros desconhecidos como quantidades aleatórias, enquanto a segunda os considera constantes. Segundo a abordagem Bayesiana, antes de qualquer dado ser coletado deve-se assumir uma distribuição de probabilidade para os parâmetros desconhecidos do modelo, chamada de distribuição a priori, a qual deve representar toda a informação existente sobre estes parâmetros. Após os dados serem observados, a informação disponível é combinada, através do teorema de Bayes, com a informação proveniente dos dados, que é representada pela função de verossimilhança associada à amostra observada. Esta combinação resulta na distribuição a posteriori dos parâmetros desconhecidos que proverá toda a informação sobre estes após os dados terem sido observados. Assim, toda a inferência desejada será retirada da posteriori. Quando não há informação disponível sobre alguns dos parâmetros, pode-se utilizar prioris não-informativas para os mesmos, e então os resultados da análise Bayesiana serão comparáveis aos resultados da análise clássica.

Para derivar analiticamente a distribuição a posteriori, alguns cálculos são necessários, mas nem sempre são exequíveis. Os métodos MCMC (acrônimo de “Markov Chain Monte Carlo”) são métodos computacionais de simulação que permitem que sejam obtidas observações que, após certo número de iterações, podem ser consideradas como provenientes da posteriori de interesse. Assim, pode-se estimar momentos, quantis e até mesmo a densidade da distribuição. Um dos métodos MCMC é o algoritmo Amostrador de Gibbs (“Gibbs Sampler”), que neste trabalho será utilizado para analisar modelos mistos aplicados a dados longitudinais, pois este modelo não possui posteriori analiticamente derivável.

Na análise de dados longitudinais, as abordagens Bayesiana e clássica podem levar a resultados diferentes, e assim, o pesquisador pode ficar em dúvida sobre qual método utilizar. Na literatura, há poucos estudos que comparam as duas abordagens para dados longitudinais sob um enfoque analítico. A maioria das discussões existentes aponta vantagens e desvantagens de cada um dos métodos apenas sob o

ponto de vista conceitual. Alguns dos argumentos a favor da abordagem Bayesiana são: a) na maioria dos problemas, o número de suposições que devem ser feitas é menor do que na análise clássica. Por exemplo, na inferência clássica para os modelos mistos, deve-se assumir que as variâncias populacionais são conhecidas para poder obter uma expressão para a variância dos estimadores, uma suposição que é raramente atendida. A consequência desse fato é um vício nos erros padrões dos efeitos fixos e aleatórios. Isto já não ocorre na análise Bayesiana, pois não é necessário supor as variâncias conhecidas na análise; b) na abordagem clássica para modelos mistos, os parâmetros são estimados por algoritmos iterativos que podem ter problemas com amostras pequenas, muitos dados faltantes, etc. A inferência Bayesiana geralmente se comporta melhor frente a essas situações; e, c) no caso dos parâmetros de variâncias do modelo misto, às vezes a análise clássica produz estimativas negativas de variâncias (método da ANOVA) ou, então, o método iterativo trunca e não fornece nenhuma estimativa (métodos ML e REML). Na inferência Bayesiana, geralmente esses problemas não ocorrem. Uma desvantagem da abordagem Bayesiana é a sua implementação computacional. O software gratuito WinBUGS já está com uma interface mais amigável do que sua primeira versão, mas ainda exige algum entendimento de programação do usuário.

Estudos de simulação comparando as inferências clássica e Bayesiana foram realizados para alguns casos especiais do modelo misto e também do modelo misto generalizado, mas não abordaram modelos para dados longitudinais. Há também estudos que abordaram o caso dos dados longitudinais, mas apenas compararam os estimadores Bayesianos entre si, sem considerar o estimador clássico.

Por isso, o objetivo deste trabalho é realizar um estudo de simulação de dados longitudinais para comparar as abordagens Bayesiana e clássica e avaliar as propriedades dos estimadores quanto ao vício e precisão.

## 2 REVISÃO DA LITERATURA

Como já mencionado, o intuito deste trabalho é comparar os métodos Bayesiano e clássico na análise de dados longitudinais, através de simulação computacional. Foram encontrados poucos estudos com objetivos similares, os quais são descritos a seguir.

Leotti e Ziegelmann (1) apresentam exemplos da aplicação do Amostrador de Gibbs ao modelos mistos e um estudo de simulação de dados longitudinais foi realizado para comparar os estimadores Bayesianos média, moda estimada não-parametricamente e mediana. Porém, este estudo não contemplou a comparação com os estimadores clássicos. O modelo simulado considerou a estrutura de covariância Componentes de Variância para ambas as matrizes de covariância dos efeitos aleatórios e do erro. Além disso, considerou a situação de simular dados com uma estrutura Não-estruturada (UN) para a matriz de covariância do erro, e analisar estes dados supondo que a estrutura da mesma era Componentes de Variância. As conclusões do estudo foram as de que, ao supor erros independentes (ou seja, estrutura Componentes de Variância para o erro), quando na verdade os dados tinham erros correlacionados e com variâncias diferentes (ou seja, provenientes de uma estrutura de covariância Não-Estruturada para o erro), os estimadores dos efeitos fixos, e dos componentes de variância se tornam viesados. Conclui também que a moda estimada não-parametricamente é um estimador com boas propriedades no caso do modelo estudado.

Browne (2) e Browne e Draper (3;4) procederam a estudos de simulação para alguns casos especiais do modelo misto e também do modelo misto generalizado, mas não abordaram modelos para dados longitudinais. Um dos modelos estudados por estes autores foi o modelo de componentes de variância (2;3), que nada mais é que um modelo de ANOVA de um fator aleatório. Os autores compararam os estimadores clássicos através dos métodos de mínimos quadrados generalizados iterativos (IGLS, da expressão em inglês), que produz resultados idênticos ao método ML, e de mínimos quadrados generalizados iterativos restritos (RIGLS, também do inglês), que é equivalente ao REML. Compararam também os estimadores

Bayesianos média, moda e mediana através de duas prioris não-informativas para os componentes de variância: priori inversa-gama e priori uniforme para a variância. No estudo foram consideradas 8 situações de tamanho de amostra e desbalanceamento, ou seja, 8 combinações de número de níveis do fator e número de repetições para cada nível. Diferentes valores para os componentes de variância também foram considerados. Nas simulações foram realizadas 1000 replicações, e para o Amostrador de Gibbs usaram um período de aquecimento de 500 e um número subsequente de iterações variou entre 10000 e 50000 dependendo da configuração.

Dentre as principais resultados, os autores destacaram que todos os métodos apresentaram viés praticamente nulo para estimar o único efeito fixo do modelo e a variância do erro aleatório. Ainda, para todos os métodos o viés diminui como aumento do número total de observações, e é menor nas situações balanceadas que não-balanceadas. O método ML subestimou a variância do efeito aleatório para todas as configurações, o que não ocorreu com o REML. Quando a priori Gama é utilizada, o estimador com menor vício para a variância do efeito aleatório é a mediana, enquanto que para a priori uniforme, o estimador menos viesado é a moda. Os autores concluíram que para o modelo de componentes de variância os métodos REML e Bayesiano são equivalentes, mas o primeiro tem a vantagem de ser mais automático.

Outro modelo estudado foi um modelo multinível (hierárquico) de 2 níveis, que os autores chamaram de modelo de regressão de coeficientes aleatórios (2;4). Toda a simulação foi baseada em um problema real em que se tinham escolas e alunos dentro das escolas, e se queria relacionar notas dos alunos em um teste de matemática em dois determinados anos. Assim, o modelo ajustava uma reta para descrever essa relação, mas cada escola tinha sua particular reta, pois se considerou escola como um efeito aleatório. Esse modelo é idêntico ao modelo de coeficientes aleatórios para dados longitudinais, pois as escolas podem ser comparadas com os indivíduos, e os alunos dentro das escolas podem ser comparados com as observações para um mesmo indivíduo. Entretanto, esse modelo não tem covariáveis, como usualmente ocorre em problemas de dados longitudinais.

Os autores novamente compararam os estimadores clássicos ML e REML e os estimadores Bayesianos através de diferentes prioris não-informativas: prioris gama e uniforme para a variância do erro aleatório, e como admitiram efeitos aleatórios correlacionados, usaram priori Wishart para a matriz de covariância dos mesmos, comparando 3 especificações desta. Consideraram 4 situações de tamanho de amostra e desbalanceamento. Foram realizadas 1000 replicações da simulação, e para o Amostrador de Gibbs usaram um período de aquecimento de 500 e um número subsequente de iterações entre 10000 e 30000 dependendo da configuração.

Os resultados das simulações revelaram que houve pouca diferença entre as duas prioris para a variância do erro e que todos os métodos apresentaram viés praticamente nulo para estimar os efeitos fixos e a variância do erro. O desbalanceamento não afeta tanto o viés quanto um número pequeno de níveis do fator aleatório (no caso, as escolas). O método ML subestimou as variâncias dos coeficientes aleatórios para todas as configurações, o que não ocorreu com o REML. Por fim, os autores não fizeram recomendação sobre qual método utilizar quando o número de níveis do efeito aleatório é pequeno, mas quando este número é moderado ou grande, recomendam que se utilize o método REML nos estágios exploratórios de uma análise, devido ao tempo computacional, e então utilizar a abordagem Bayesiana para obter as estimativas do modelo final.

Portanto, um estudo de simulação que compare os métodos Bayesiano e clássico na análise de dados longitudinais ainda não foi publicado. Para proceder tal estudo, foi necessário revisar a literatura dos métodos MCMC e das abordagens clássica e Bayesiana para os modelos mistos. Esta revisão está apresentada na seqüência.

## **2.1 MÉTODOS MCMC**

Atualmente, muitos problemas estatísticos são da categoria de modelos de grande dimensão, pois contêm muitos parâmetros. Como exemplo, Gamerman (5) cita os modelos para dados espaciais. Até recentemente, trabalhar com a estrutura complexa destes modelos em muitas aplicações era difícil, pois exigia o

desenvolvimento de métodos e rotinas computacionais específicos para cada caso (6). Os métodos envolvendo simulação via MCMC fornecem uma abordagem única para analisar muitos destes problemas.

Os métodos Monte Carlo são úteis quando a partir de uma variável aleatória  $X$  com distribuição conhecida consegue-se, através de algum algoritmo, gerar valores para estudar o comportamento da distribuição de uma função  $g_X(\cdot)$ , que pode ser impossível de ser trabalhada analiticamente. Assim, a simulação de Monte Carlo consiste basicamente em gerar valores independentes de  $X$ , calcular  $g(x_i)$  correspondente a cada valor  $x_i$  simulado e, então, avaliar o comportamento aproximado de características da distribuição de  $g_X(\cdot)$ . Utilizando-se MCMC, apesar de não se conseguir gerar observações diretamente da distribuição de interesse, geram-se amostras a partir de cadeias de Markov construídas de tal forma que, após um grande número de iterações, a distribuição das observações geradas se aproxima da distribuição desejada. Com a amostra construída, as características da distribuição de interesse são aproximadas pelas respectivas características amostrais. As cadeias de Markov serão definidas e descritas na seção 2.1.1.

Duas grandes vantagens dos métodos MCMC são: a) não é necessário conhecer a forma da distribuição de que se quer simular, o que ocorre em muitos problemas práticos na inferência Bayesiana; e, b) há muitos algoritmos para se construir as cadeias de Markov necessárias para a simulação, sendo que todos atingem o objetivo de gerar dados da distribuição de interesse. Todos os algoritmos são casos particulares do algoritmo Metropolis-Hastings, que será descrito na seção 2.1.2. Um destes casos é o Amostrador de Gibbs, que será abordado na seção 2.1.3.

Gilks et al. (6) comentam que foram necessários cerca de 40 anos para o MCMC penetrar nos principais problemas práticos da Estatística. Surgiu originalmente em estudos de física-estatística, e tem sido muito utilizado em estatística espacial e análise de imagens. Nos últimos anos, houve um grande crescimento no número de publicações envolvendo aplicações dos métodos MCMC na inferência Bayesiana e também na clássica.

### 2.1.1 Cadeias de Markov

Esta seção apresenta brevemente conceitos básicos sobre cadeias de Markov (7). Uma cadeia de Markov é um tipo especial de processo estocástico, definido como uma coleção de variáveis aleatórias  $\{\theta^{(t)} : t \in T\}$  para algum conjunto  $T$ , que pode ser discreto ou contínuo. Neste trabalho será considerado somente o caso discreto, mais especificamente onde  $T = \mathbf{N}$ , e assim pode-se pensar o índice das variáveis aleatórias como o índice da iteração de uma simulação. O espaço dos possíveis valores de  $\theta^{(t)}$  é denominado espaço de estados e denotado por  $S$ .

A principal característica de uma cadeia de Markov é que, a cada tempo  $t \geq 0$ , o próximo estado da cadeia,  $\theta^{(t+1)}$ , depende do passado somente através do estado atual  $\theta^{(t)}$ . Isto pode ser formalmente definido como:

$$P(\theta^{(t+1)} = y | \theta^{(t)} = x_t, \theta^{(t-1)} = x_{t-1}, \dots, \theta^{(t-k)} = x_{t-k}, \dots, \theta^{(0)} = x_0) = P(\theta^{(t+1)} = y | \theta^{(t)} = x_t) \quad (2.1)$$

para todos  $x_0, \dots, x_t, y \in S$ ,  $k \in \mathbf{N}$ .

Em geral, as probabilidades definidas na equação (2.1) dependem de  $x_t, y$  e  $t$ . Quando não dependem de  $t$ , a cadeia é chamada de homogênea. Neste caso, a equação (2.1) é denotada por  $P_{xy}$  e chamada de probabilidade de transição estacionária a um passo. Estas probabilidades satisfazem:

$$\text{a) } P_{xy} \geq 0, \forall x, y \in S;$$

$$\text{b) } \sum_{y \in S} P_{xy} = 1, \forall x \in S.$$

Define-se também  $P(\theta^{(t+1)} | \theta^{(t)})$  como função, ou *kernel*, de transição da cadeia.

Um problema fundamental das cadeias de Markov no contexto da simulação é o estudo do comportamento assintótico da cadeia quando o número de iterações  $t \rightarrow \infty$ . Para entender este comportamento, é necessário apresentar o conceito de distribuição estacionária, denotada por  $\pi$ . A distribuição  $\pi$  é uma distribuição estacionária de uma cadeia com probabilidades de transição  $P_{xy}$  se e somente se

$$\sum_{x \in S} \pi_x P_{xy} = \pi_y, \forall y \in S.$$

Esta distribuição também é conhecida como distribuição invariante ou de equilíbrio.

Sob certas condições de regularidade (7), a probabilidade de transição a  $t$  passos, denotada por  $P_{xy}^{(t)}$ , está relacionada com a distribuição estacionária através de

$$\lim_{t \rightarrow \infty} P_{xy}^{(t)} = \pi_y. \quad (2.2)$$

Isto significa que, se o estado inicial da cadeia é  $\theta^{(0)} = x$ , então a cadeia irá gradualmente “esquecer” seu estado inicial e  $P_{xy}^{(t)}$  converge para a distribuição estacionária, que é única (7). Denota-se por  $P_x^{(t)}$  a distribuição de probabilidade que fornece as probabilidades com que a cadeia sai de seu ponto inicial  $x$  e visita qualquer outro estado do espaço de estados, em  $t$  passos. Assim, à medida que  $t$  aumenta, os pontos  $\theta^{(t)}$  gerados a partir da função de transição  $P(\theta^{(t+1)} | \theta^{(t)})$  irão cada vez mais se parecer com amostras da distribuição estacionária  $\pi$ .

Logo, a questão essencial da simulação por MCMC é definir qual função de transição utilizar para que a distribuição estacionária  $\pi$  seja aquela de interesse. Como foi mencionado, há muitas maneiras de se escolher a função de transição para obter amostras de  $\pi$ , porém, todos eles são casos especiais do algoritmo de Metropolis–Hastings, que será visto a seguir.

### 2.1.2 Algoritmo de Metropolis-Hastings

Hastings (8) generalizou o método inicialmente proposto por Metropolis et al. (9), originando assim o algoritmo de Metropolis-Hastings. A idéia do algoritmo é que, a cada tempo  $t$ , gera-se um ponto candidato  $\theta^C$  a partir de uma distribuição proposta  $Q(\cdot | \theta^{(t)})$  e, então, se aceita que o próximo estado da cadeia,  $\theta^{(t+1)}$ , seja igual a  $\theta^C$  com probabilidade  $\alpha(\theta^{(t)}, \theta^C)$ , onde

$$\alpha(\theta^{(t)}, \theta^c) = \min\left(1, \frac{\pi(\theta^c)Q(\theta^{(t)} | \theta^c)}{\pi(\theta^{(t)})Q(\theta^c | \theta^{(t)})}\right). \quad (2.3)$$

Se o ponto candidato não for aceito, a cadeia não se move, isto é,  $\theta^{(t+1)} = \theta^{(t)}$ . Portanto, o algoritmo é definido formalmente como:

1. Definir  $t = 0$  e fornecer um valor arbitrário para  $\theta^{(0)}$ ;
2. Gerar um valor  $\theta^c$  a partir de  $Q(\cdot | \theta^{(t)})$ ;
3. Gerar um valor  $U$  de uma distribuição *Uniforme*(0,1);
4. Se  $U \leq \alpha(\theta^{(t)}, \theta^c)$ , fazer  $\theta^{(t+1)} = \theta^c$ , caso contrário, fazer  $\theta^{(t+1)} = \theta^{(t)}$ ;
5. Incrementar  $t$ ;
6. Voltar ao passo 2 e continuar as iterações até um tamanho  $m$  suficientemente grande.

Sob certas condições de regularidade, a distribuição estacionária da cadeia será  $\pi(\cdot)$ , independente da forma da distribuição proposta  $Q(\cdot | \cdot)$  (6). Entretanto, ao analisar a expressão em (2.3), percebe-se que é necessário conhecer  $\pi(\cdot)$  para implementar o algoritmo e, assim, parece não ser muito útil.

Em problemas Bayesianos,  $\pi(\cdot)$  será a distribuição a posteriori dos parâmetros, que é dada por

$$\pi(\cdot) = \pi(\theta | x) = \frac{\pi(\theta)L(\theta | x)}{f(x)}, \quad (2.4)$$

onde  $\pi(\theta)$  é a distribuição a priori, que sempre será conhecida,  $L(\theta | x)$  é a função de verossimilhança de  $\theta$  para os dados observados, que também sempre terá uma forma suposta, e  $f(x) = \int \pi(\theta)L(\theta | x)d\theta$  que pode não ter forma analítica fechada em alguns casos. Note, porém, que  $f(x)$  é só uma constante de normalização, ou seja, ela nunca envolverá  $\theta$  em sua expressão. Por isto, substituindo (2.4) em (2.3), produz:

$$\frac{\pi(\theta^c)Q(\theta^{(t)}|\theta^c)}{\pi(\theta^{(t)})Q(\theta^c|\theta^{(t)})} = \frac{\pi(\theta^c|x)Q(\theta^{(t)}|\theta^c)}{\pi(\theta^{(t)}|x)Q(\theta^c|\theta^{(t)})} = \dots = \frac{\pi(\theta^c)L(\theta^c|x)Q(\theta^{(t)}|\theta^c)}{\pi(\theta^{(t)})L(\theta^{(t)}|x)Q(\theta^c|\theta^{(t)})} \quad (2.5)$$

que só envolve expressões conhecidas. Ou seja, não é preciso conhecer a distribuição a posteriori para aplicar o algoritmo Metropolis-Hastings.

A seguir será apresentada a técnica do Amostrador de Gibbs, também conhecida na Língua Inglesa como “Gibbs Sampler” e que é adequada quando o modelo possui mais de um parâmetro desconhecido. Consiste basicamente no algoritmo de Metropolis-Hastings onde a distribuição proposta para cada parâmetro a ser simulado é a respectiva distribuição condicional completa e a probabilidade de aceitação de cada valor gerado será sempre igual a um.

### 2.1.3 Amostrador de Gibbs

A origem do Amostrador de Gibbs está no artigo de Geman e Geman (10) que trata de processamento de imagens. Neste artigo, os autores utilizaram o algoritmo para simular dados de uma distribuição de Gibbs, por isto a denominação de Amostrador de Gibbs. Entretanto, somente com o trabalho de Gelfand e Smith (11), a técnica encontrou aplicações em vários problemas estatísticos além da análise de imagens.

#### 2.1.3.1 Definição e propriedades

Seja  $\boldsymbol{\theta}$  um vetor  $p \times 1$ , ou seja,  $\boldsymbol{\theta}' = (\theta_1, \dots, \theta_p)$ , onde cada componente  $\theta_i$ ,  $i = 1, \dots, p$ , pode ser um escalar, um vetor ou uma matriz. O interesse está na distribuição  $\pi(\boldsymbol{\theta})$  que, nos problemas práticos, pode não ter uma densidade de probabilidade analiticamente derivável ou possuir um algoritmo de geração direta muito trabalhoso ou simplesmente inexistente. Se as distribuições  $\pi(\theta_i | \boldsymbol{\theta}_{-i})$ ,  $i = 1, \dots, p$  e  $\boldsymbol{\theta}'_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ , denominadas distribuições condicionais

completas, são totalmente conhecidas e é possível simular observações a partir delas, então se pode aplicar o Amostrador de Gibbs através do seguinte algoritmo:

1. Atribuir valores iniciais para  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$  e definir  $t = 0$ ;
2. Obter o próximo estado da cadeia,  $\boldsymbol{\theta}^{(t+1)} = (\theta_1^{(t+1)}, \dots, \theta_p^{(t+1)})$ , a partir das seguintes gerações sucessivas:

$$\begin{aligned}\theta_1^{(t+1)} &\sim \pi(\theta_1 | \theta_2^{(t)}, \dots, \theta_p^{(t)}) \\ \theta_2^{(t+1)} &\sim \pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}) \\ &\vdots \\ \theta_p^{(t+1)} &\sim \pi(\theta_p | \theta_1^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})\end{aligned}$$

3. Incrementar  $t$ ;
4. Voltar ao passo 2 e continuar as iterações até um tamanho  $m$  que garanta a convergência do algoritmo (descrito na próxima seção).

Como já mencionado, o Amostrador de Gibbs é um caso particular do algoritmo de Metropolis-Hastings, com distribuições propostas iguais às condicionais completas. Como a probabilidade de aceitação de cada valor gerado através do Amostrador de Gibbs é igual a 1, a cadeia sempre se moverá (6). O Amostrador de Gibbs reduz a dimensão do problema, pois ao invés de gerar amostras do vetor  $\boldsymbol{\theta}$  inteiro, gera valores para blocos de componentes de  $\boldsymbol{\theta}$  de menores dimensões (6).

Este algoritmo define uma cadeia de Markov, já que cada novo estado da cadeia,  $\boldsymbol{\theta}^{(t+1)}$ , depende somente dos valores obtidos na iteração  $t$ , e a distribuição estacionária da cadeia de Markov definida pelo Amostrador de Gibbs é  $\pi$ . Assim, quando  $t \rightarrow \infty$ , os valores simulados para  $\boldsymbol{\theta}$  são amostras de  $\pi(\boldsymbol{\theta})$  e as observações geradas para cada  $\theta_i$ ,  $i = 1, \dots, p$ , são amostras de  $\pi(\theta_i)$ . Uma prova formal da convergência do algoritmo, bem como as condições para o seu acontecimento, pode ser encontrada em Tierney (12).

### 2.1.3.2 Geração da amostra, valores iniciais e convergência

Como o Amostrador de Gibbs define uma cadeia de Markov, a amostra gerada através dele não é independente. Cada novo valor depende dos valores gerados na iteração anterior. Por este motivo, há duas maneiras de se obter uma amostra aproximadamente independente de tamanho  $n$  da distribuição de interesse  $\pi$ . A primeira forma é replicar o algoritmo  $n$  vezes, em cada uma delas fazendo  $m$  iterações e então tomar o  $m$ -ésimo valor gerado de  $\theta$  como um dos elementos da amostra de tamanho  $n$ . Esta simulação precisará de  $mn$  gerações da cadeia. Se forem estabelecidos valores para os estados iniciais da cadeia  $\theta_j^{(0)}$ ,  $j=1, \dots, n$ , de forma independente, a amostra consiste em valores independentes de  $\pi(\theta)$  (5).

Outra forma é considerar uma única replicação do algoritmo, baseando-se no fato de que, após a convergência, os valores simulados da cadeia seguem a distribuição estacionária  $\pi(\theta)$ . Assim, os  $b$  primeiros valores gerados são considerados como um período de aquecimento do algoritmo (da expressão em inglês “burn-in”), necessário para a convergência do Amostrador de Gibbs, e são descartados, utilizando-se somente os  $n$  valores posteriores como uma amostra da distribuição de interesse. Para diminuir a dependência entre os valores simulados, pode-se determinar um valor  $k$ , conhecido como “thin” em inglês, pelo fato deste procedimento “emagrecer” a amostra. A amostra será formada pelos valores gerados da cadeia a cada  $k$ -ésima iteração após o período de aquecimento, ou seja,  $\theta_i^{(b+1)}, \theta_i^{(b+k+1)}, \theta_i^{(b+2k+1)}, \dots$ . Entretanto, não se ganha em eficiência utilizando este procedimento e a estimação é sempre menos precisa do que se mantendo na amostra todos os valores gerados após o período de aquecimento (5).

Gilks et al. (6) e Gamerman (5) comentam que há conflitos na literatura sobre qual abordagem utilizar. Porém, ambos não recomendam a geração de  $n$  cadeias independentes, pois o custo computacional é maior, enquanto uma cadeia mais longa diminuiria a dependência do valor inicial.

Sobre o valor inicial,  $\theta^{(0)}$ , a ser estabelecido, não há muito material escrito, já que, teoricamente, ele não afeta a distribuição estacionária da cadeia. Devem ser

escolhidos com mais cuidado para cadeias de convergência mais lenta, para se evitar um longo período de aquecimento (6).

O estabelecimento do tamanho do período de aquecimento não pode ser realizado analiticamente na maioria dos casos. Entretanto, existem métodos informais e formais para se verificar a convergência da cadeia. O método informal mais utilizado baseia-se em gráficos da trajetória da cadeia, para verificar a partir de qual iteração os dados passam a apresentar um mesmo padrão de comportamento. Um exemplo é dado no Anexo 3. Os métodos formais, conhecidos como diagnósticos de convergência, usam uma variedade de resultados teóricos e aproximações, mas todos utilizam os valores simulados de alguma maneira. Cowles e Carlin (13) e Brooks e Roberts (14) apresentam uma revisão dos principais métodos.

Existem também algoritmos de diagnóstico de convergência, como o CODA, anagrama para a abreviatura da expressão “Convergence Diagnostics and Output Analysis” (15). Este programa está disponível como um pacote para o software R e inserido no WinBUGS. Este último é um software específico para a implementação do Amostrador de Gibbs. A expressão BUGS é um acrônimo de “Bayesian inference Using Gibbs Sampling”, cuja primeira versão foi apresentada por Spiegelhalter et al. (16) e é distribuído gratuitamente na internet, assim como o R.

### 2.1.3.3 Análise dos resultados da simulação

Considerando-se uma simulação por Amostrador de Gibbs através do algoritmo descrito na seção 2.1.3.1, com uma única cadeia de tamanho  $m$  e um período de aquecimento de tamanho  $b$ , a média e a variância de  $\theta_i$  podem ser estimados como:

$$\bar{\theta}_i = \frac{1}{m-b} \sum_{t=b+1}^m \theta_i^{(t)} \quad (2.6)$$

$$e \ s_i^2 = \frac{1}{m-b-1} \sum_{t=b+1}^m (\theta_i^{(t)} - \bar{\theta}_i)^2 . \quad (2.7)$$

Ainda, qualquer quantil da distribuição, como, por exemplo, a mediana, é estimado obtendo-se o correspondente quantil na amostra constituída pelos valores gerados após o período de aquecimento. A própria densidade marginal de  $\theta_i$  pode ser estimada através da equação

$$\hat{\pi}(\theta_i) = \frac{1}{m-b} \sum_{t=b+1}^m \pi(\theta_i | \boldsymbol{\theta}_{-i}^{(t)}). \quad (2.8)$$

Este estimador para  $\pi(\theta_i)$  é baseado na esperança da densidade condicional

$$E[\pi(\theta_i | \boldsymbol{\theta}_{-i})] = \int \pi(\theta_i | \boldsymbol{\theta}_{-i}) \pi(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i} = \pi(\theta_i). \quad (2.9)$$

Logo, o estimador proposto por (2.8) imita (2.9), já que  $\boldsymbol{\theta}_{-i}^{(t)}$ ,  $t = b+1, \dots, m$  são considerados uma amostra de  $\pi(\boldsymbol{\theta}_{-i})$  (17). Deve-se salientar que a expressão em (2.8) estima o valor da densidade de probabilidade  $\pi(\theta_i)$  para um específico valor de  $\theta_i$ . Para se obter uma estimativa da densidade de toda a distribuição, pode-se arbitrar um conjunto de valores representativos do espaço amostral de  $\theta_i$  e avaliar a expressão em (2.8) para estes valores. Outra opção é estimar a densidade através da suavização de Kernel (“Kernel-Smoothing”) a partir dos valores observados. Este procedimento é realizado pelo software WinBUGS e foi exemplificado no Anexo 3.

Quando  $\theta_i$  é contínuo, a moda da sua distribuição não pode ser estimada através do valor mais freqüente dentre os dados gerados a partir do período de aquecimento. Isto acontece porque, ao se gerar números aleatórios com várias casas decimais, dificilmente serão obtidos valores idênticos. Como alternativa, pode-se calcular a densidade marginal estimada por (2.8) e observar o valor de  $\theta_i$  que a maximiza. Outro método possível, computacionalmente mais rápido que a estimativa da densidade através de (2.8), seria observar o valor que maximiza a densidade marginal de  $\theta_i$  estimada por suavização de Kernel.

Para uma função real de  $\theta_i$ , denotada por  $\psi = t(\theta_i)$ , os estimadores para média e variância são análogos aos apresentados em (2.6) e (2.7). Ou seja,

$$\hat{E}(\psi) = \bar{\psi} = \frac{1}{m-b} \sum_{t=b+1}^m \psi_t = \frac{1}{m-b} \sum_{t=b+1}^m t(\theta_i^{(t)})$$

$$\text{e } s_{\psi}^2 = \frac{1}{m-b-1} \sum_{t=b+1}^m (\psi_t - \bar{\psi})^2 .$$

## 2.2 ABORDAGEM CLÁSSICA PARA MODELOS MISTOS NORMAIS

Para muitas aplicações estatísticas, o uso do modelo linear geral (GLM) é uma metodologia muito útil. Este modelo supõe que os termos de erro são normalmente distribuídos, independentes e com variância constante. Dois modelos conhecidos que podem ser escritos desta forma são os modelos de análise de variância e de análise de regressão. Entretanto, há casos em que a suposição de independência dos erros é muito restritiva e não pode ser considerada válida. Um exemplo é a análise de dados longitudinais, onde se tomam observações em uma mesma unidade de observação ao longo do tempo e, por este fato, provavelmente haverá correlação entre as observações realizadas no mesmo indivíduo. Assim, os modelos mistos permitem analisar estes casos, já que se podem assumir erros correlacionados e com variâncias heterogêneas. Além disto, é possível modelar efeitos fixos e aleatórios, enquanto o modelo linear geral somente modela efeitos fixos. Assim como no GLM, o modelo misto também pode ser utilizado nos casos de dados não balanceados. Apesar de todas estas características favoráveis dos modelos mistos, somente com o maior desenvolvimento de softwares como SAS, MLwiN e R, entre outros, a metodologia de modelos mistos se tornou mais acessível. Os modelos mistos normais, que serão abordados neste trabalho, são aqueles em que se assume distribuição normal para os erros e efeitos aleatórios.

Henderson (18) aponta como maior vantagem do modelo misto a unificação da análise de várias técnicas estatísticas, permitindo que se utilizem os mesmos métodos de estimação, testes de hipóteses e predição para muitos problemas práticos.

A seção 2.2.1 apresenta a formulação do modelo misto. Na seqüência, a seção 2.2.2 apresenta os métodos de estimação dos efeitos fixos, aleatórios e parâmetros de

variâncias e covariâncias. A seção 2.2.3 trata especificamente de modelos mistos aplicados ao caso de dados longitudinais.

### 2.2.1 Formulação do modelo misto

Em notação matricial, pode-se expressar o modelo misto como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (2.10)$$

onde:

$\mathbf{y}$  é um vetor  $n \times 1$  de dados observados;

$\mathbf{X}$  é uma matriz de planejamento  $n \times p$  conhecida;

$\boldsymbol{\beta}$  é um vetor  $p \times 1$  de efeitos fixos geralmente desconhecidos;

$\mathbf{Z}$  é uma matriz de planejamento ou de incidência  $n \times q$  conhecida;

$\mathbf{u}$  é um vetor  $q \times 1$  de efeitos aleatórios geralmente desconhecidos; e,

$\boldsymbol{\varepsilon}$  é um vetor  $n \times 1$  de erros aleatórios não-observáveis.

As suposições do modelo são:  $\mathbf{u} \sim N(0, \mathbf{D})$ ,  $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R})$  e  $Cov(\mathbf{u}, \boldsymbol{\varepsilon}') = 0$ , onde  $\mathbf{D}$  e  $\mathbf{R}$  são matrizes de variâncias e covariâncias de dimensões  $q \times q$  e  $n \times n$ , respectivamente. Nota-se, por (2.10), que o modelo linear geral é um caso especial de modelo misto onde  $\mathbf{Z}$  é nula e  $\mathbf{R} = \mathbf{I}_n \sigma^2$ , sendo  $\mathbf{I}_n$  a matriz identidade  $n \times n$ .

Pela formulação do modelo em (2.10) e as suposições citadas acima, segue que  $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ ,  $Var(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} = \mathbf{V}$  e  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ .

Na análise de modelos mistos, pode-se supor diferentes padrões para as matrizes de variâncias e covariâncias  $\mathbf{D}$  e  $\mathbf{R}$ , modelando uma estrutura de covariância para elas. Existem vários tipos de estruturas de covariância que podem ser modeladas, tais como:

1. *Componentes de Variância* (VC – “Variance Components”): permite modelar variâncias diferentes para grupos de efeitos ou observações e supõe covariâncias nulas.

$$VC = \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \ddots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma_A^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma_B^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \ddots & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \sigma_B^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

2. *Autoregressiva de 1ª ordem* (AR(1) – “Autoregressive (1)”): supõe variâncias iguais, mas as covariâncias diminuem à medida que os elementos se afastam da diagonal principal da matriz.

$$AR(1) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots \\ \rho & 1 & \rho & \rho^2 & \dots \\ \rho^2 & \rho & 1 & \rho & \dots \\ \rho^3 & \rho^2 & \rho & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

3. *Não-estruturada* (UN – “Unstructured”): é o caso mais geral possível de estrutura de covariância, permitindo variâncias e covariâncias diferentes.

$$UN = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} & \dots \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} & \dots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} & \dots \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Pode-se especificar o modelo misto em (2.10) na forma de um modelo hierárquico (19;20). Modelos hierárquicos, em linhas gerais, são modelos especificados em estágios, onde cada estágio possui um submodelo. São muito úteis, assim como os modelos mistos, para dados que se apresentam agrupados em níveis hierárquicos onde cada estágio modela as relações entre as variáveis daquele nível. Entretanto, apesar da formulação desses modelos serem diferentes, elas são equivalentes. No caso dos modelos mistos, tem-se que:

$$\text{Est\u00e1gio 1: } \mathbf{y} | \mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}; \mathbf{R}) \quad (2.11)$$

$$\text{Est\u00e1gio 2: } \mathbf{u} \sim N(\mathbf{0}; \mathbf{D}). \quad (2.12)$$

Bryk e Raudenbush (21) denominam o modelo acima de modelo linear hier\u00e1rquico condicional. Esta hierarquia ser\u00e1 \u00fasil para se compreender algumas propriedades do modelo misto, bem como a aplica\u00e7\u00e3o das t\u00e9cnicas Bayesianas a este modelo, assunto que ser\u00e1 abordado na se\u00e7\u00e3o 2.3.

### 2.2.2 Estima\u00e7\u00e3o

Na infer\u00eancia cl\u00e1ssica, os m\u00e9todos de estima\u00e7\u00e3o usualmente baseiam-se na maximiza\u00e7\u00e3o da fun\u00e7\u00e3o de verossimilhan\u00e7a, segundo a id\u00e9ia de que a estimativa mais adequada de um par\u00e2metro \u00e9 o valor mais prov\u00e1vel de ter gerado o particular conjunto de dados observados.

Em modelos em que se sup\u00f5em observa\u00e7\u00f5es independentes, a verossimilhan\u00e7a \u00e9 simplesmente o produto da fun\u00e7\u00e3o densidade de probabilidade de cada observa\u00e7\u00e3o. Como no modelo misto as observa\u00e7\u00f5es nem sempre s\u00e3o independentes, a verossimilhan\u00e7a precisa ser baseada em uma distribui\u00e7\u00e3o multivariada. Assim, para o modelo misto normal, a fun\u00e7\u00e3o de verossimilhan\u00e7a \u00e9 baseada em uma distribui\u00e7\u00e3o normal multivariada, como segue:

$$L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{V}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \quad (2.13)$$

onde  $|\mathbf{V}|$  \u00e9 o determinante da matriz  $\mathbf{V}$ .

Para a maximiza\u00e7\u00e3o da fun\u00e7\u00e3o de verossimilhan\u00e7a, \u00e9 mais f\u00e1cil utilizar o seu logaritmo natural, como abaixo:

$$\ln L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y}) = -\frac{1}{2} n \ln(2\pi) - \frac{1}{2} [\ln|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]. \quad (2.14)$$

Há interesse em se fazer inferências sobre três tipos de parâmetros quando se trabalha com modelos mistos: os efeitos fixos, os efeitos aleatórios e os componentes de variância do modelo (variâncias e covariâncias das matrizes  $\mathbf{D}$  e  $\mathbf{R}$ ).

### 2.2.2.1 Estimação dos efeitos fixos e aleatórios

Há três maneiras de se obter uma solução para os efeitos fixos: a primeira delas consiste em obter o valor de  $\boldsymbol{\beta}$  que maximiza (2.14); a segunda é obter uma solução através de mínimos quadrados generalizados e, a terceira, é solucionar as equações do modelo misto. Será mostrado a seguir que as três maneiras obtêm o mesmo estimador para  $\boldsymbol{\beta}$ . Para os efeitos aleatórios, pode-se estimá-los através das equações do modelo misto.

**Solução 1:** Para maximizar a função de verossimilhança, deve-se diferenciar (2.14) em relação à  $\boldsymbol{\beta}$  considerando-se  $\mathbf{V}$  conhecida, e igualar o resultado a zero, obtendo uma solução que depende dos parâmetros de variâncias e covariâncias do modelo:

$$\mathbf{X}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}.$$

Fazendo algumas manipulações na expressão acima, obtém-se

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad (2.15)$$

cuja esperança e variância são, respectivamente,

$$E(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\mathbf{y}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \quad (2.16)$$

e

$$Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Var(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \dots = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad (2.17)$$

O estimador definido na equação (2.15) é não-viciado quando se considera  $\mathbf{V}$  conhecida. Posteriormente, será discutido o caso em que a matriz  $\mathbf{V}$  não é conhecida.

**Solução 2:** O método de mínimos quadrados generalizados, no caso dos modelos mistos, consiste em minimizar o produto dos resíduos completos,  $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ , ponderados pelo inverso da matriz de variâncias,  $\mathbf{V}^{-1}$ . O termo de resíduos completos é devido à diferenciação entre esses e os resíduos ordinários,  $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}$ . Logo, um estimador para  $\boldsymbol{\beta}$  é obtido minimizando-se a seguinte expressão em relação à  $\boldsymbol{\beta}$ :

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.18)$$

Sem fazer cálculos, pode-se notar que o estimador obtido por mínimos quadrados generalizados é o mesmo obtido por máxima verossimilhança, já que maximizar (2.14) em relação à  $\boldsymbol{\beta}$  nada mais é do que minimizar (2.18), também em relação a  $\boldsymbol{\beta}$ .

**Solução 3:** As equações do modelo misto são devidas a Henderson (18) e são úteis para obter tanto soluções para os parâmetros dos efeitos fixos quanto para os parâmetros dos efeitos aleatórios. São obtidas através da maximização da função de verossimilhança em termos de  $\boldsymbol{\beta}$ ,  $\mathbf{u}$ ,  $\mathbf{D}$  e  $\mathbf{R}$ , construída considerando o modelo misto como um modelo hierárquico.

Assim, usando as equações definidas em (2.11) e (2.12),

$$\begin{aligned} L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) &= f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y} | \mathbf{u}) f(\mathbf{u}) = L(\boldsymbol{\beta}, \mathbf{R} | \mathbf{u}, \mathbf{y}) \times L(\mathbf{D} | \mathbf{u}) \quad (2.19) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]' \mathbf{R}^{-1} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]\right\} \times \frac{1}{(2\pi)^{\frac{q}{2}} |\mathbf{D}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \mathbf{u}' \mathbf{D}^{-1} \mathbf{u}\right\} \end{aligned}$$

Como no caso da maximização de (2.14), é mais fácil maximizar o logaritmo da equação (2.19), como segue:

$$\begin{aligned} l = \ln L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) &= \frac{n+q}{2} \ln(2\pi) - \frac{1}{2} \left\{ \ln|\mathbf{R}| + [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]' \mathbf{R}^{-1} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})] \right. \\ &\quad \left. + \ln|\mathbf{D}| + \mathbf{u}' \mathbf{D}^{-1} \mathbf{u} \right\}. \quad (2.20) \end{aligned}$$

Derivando a equação (2.20) em relação à  $\boldsymbol{\beta}$  e  $\mathbf{u}$ , e considerando  $\mathbf{D}$  e  $\mathbf{R}$  conhecidas, as substituições de  $\boldsymbol{\beta}$  por  $\hat{\boldsymbol{\beta}}$  e de  $\mathbf{u}$  por  $\hat{\mathbf{u}}$  produz

$$\begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \mathbf{u}} \end{bmatrix} = \begin{bmatrix} -\mathbf{X}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ -\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{D}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{u}} + \mathbf{D}^{-1}\hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}.$$

Com algumas manipulações, e lembrando que  $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$ , são obtidos os estimadores:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad \text{e} \quad (2.21)$$

$$\hat{\mathbf{u}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.22)$$

Portanto, o estimador para os parâmetros dos efeitos fixos em (2.21) é idêntico àquele definido em (2.15), e sua variância, por consequência, é idêntica àquela definida na equação (2.17). A variância de  $\hat{\mathbf{u}}$  é

$$\text{Var}(\hat{\mathbf{u}}) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D} - \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Z}\mathbf{D}. \quad (2.23)$$

Note, por (2.21) e (2.22), que os estimadores de  $\boldsymbol{\beta}$  e  $\mathbf{u}$  supõem que a matriz  $\mathbf{V}$  seja conhecida. Em problemas práticos, raramente a matriz  $\mathbf{V}$  será conhecida. Nestas situações,  $\boldsymbol{\beta}$  e  $\mathbf{u}$  são estimados substituindo  $\mathbf{V}$  por  $\hat{\mathbf{V}}$  (uma estimativa para  $\mathbf{V}$ ) em, respectivamente, (2.21) e (2.22). Para as estimativas dos parâmetros dos efeitos fixos e aleatórios, este procedimento não acarreta problemas, mas subestima as variâncias dos estimadores (22).

Se  $\mathbf{D}$  e  $\mathbf{R}$  são conhecidas,  $\hat{\boldsymbol{\beta}}$  é o melhor estimador linear não-viciado (BLUE – “Best Linear Unbiased Estimator”) de  $\boldsymbol{\beta}$  e  $\hat{\mathbf{u}}$  é o melhor preditor linear não-viciado (BLUP – “Best Linear Unbiased Predictor”) de  $\mathbf{u}$ , onde “melhor” significa erro quadrático médio mínimo. De qualquer modo,  $\mathbf{D}$  e  $\mathbf{R}$  são usualmente desconhecidas e estimadas utilizando um dos métodos de estimação de componentes de variâncias a serem vistos a seguir. Neste caso, as abreviaturas BLUE e BLUP não são adequadas, mas a palavra empírico é freqüentemente acrescentada para indicar tal aproximação. A abreviação apropriada se torna então EBLUE e EBLUP (23).

#### 2.2.2.2 Estimação dos componentes de variâncias

Existem vários métodos disponíveis para estimar os parâmetros de variâncias e covariâncias do modelo misto. Os três métodos mais conhecidos são: o método da ANOVA, ou dos momentos; o método da máxima verossimilhança (ML) e o método da máxima verossimilhança restrita, ou residual (REML).

O método da ANOVA consiste basicamente em igualar os quadrados médios observados aos seus valores esperados, em uma tabela de análise de variância. Note que os valores esperados são combinações lineares dos componentes de variância. Assim, obtêm-se equações lineares que envolvem os componentes de variância, e as soluções são tomadas como seus estimadores. Os estimadores produzidos pelo método da ANOVA são sempre não-viciados e de variância mínima. Porém, podem produzir estimativas negativas de componentes de variância, o que não é coerente.

O método da Máxima Verossimilhança (ML) maximiza a função de verossimilhança definida em (2.13) em relação aos parâmetros de variâncias e covariâncias, considerando os parâmetros dos efeitos fixos como constantes. Após obter as estimativas para os parâmetros de variâncias e covariâncias, os efeitos fixos são então estimados maximizando a mesma verossimilhança, porém considerando os parâmetros de variâncias e covariâncias como constantes. Este método produz estimativas viciadas (subestima) para os parâmetros de variâncias e covariâncias.

Quanto menor o número de graus de liberdade utilizado para estimar os parâmetros de variâncias e covariâncias, maior é o viés (22).

O método da Máxima Verossimilhança Residual ou Restrita (REML) utiliza uma função de verossimilhança que não envolve  $\boldsymbol{\beta}$ , ou seja, é definida somente em termos dos parâmetros de variâncias e covariâncias. O primeiro passo é obter uma função de verossimilhança a partir dos resíduos completos estimados,  $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ , onde  $\hat{\boldsymbol{\beta}}$  está definido na equação (2.15). Como  $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\beta}}$  são independentes e  $\hat{\boldsymbol{\beta}}$  segue uma distribuição normal multivariada com média e variância dadas por (2.16) e (2.17), a verossimilhança restrita é escrita como

$$L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y})}{L(\boldsymbol{\beta} | \hat{\boldsymbol{\beta}}, \mathbf{V})},$$

onde  $L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y})$  foi definida em (2.13) e

$$L(\boldsymbol{\beta} | \hat{\boldsymbol{\beta}}, \mathbf{V}) = \frac{1}{(2\pi)^{\frac{p}{2}} |(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}.$$

Portanto,

$$L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{(2\pi)^{\frac{n-p}{2}} |(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|^{\frac{1}{2}} |\mathbf{V}|^{-\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\right\},$$

cujo logaritmo natural é

$$\ln L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = -\frac{1}{2} \left\{ (n-p) \ln(2\pi) - \ln |(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}| + \ln |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right\} \quad (2.24)$$

Apesar de  $\hat{\boldsymbol{\beta}}$  aparecer em (2.24), a função de verossimilhança só depende dos parâmetros de variâncias e covariâncias, pois  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ . Deve-se notar que a diferença entre o logaritmo da verossimilhança do método ML e o logaritmo da verossimilhança do método REML está no termo adicional  $\ln |(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|$ , que é o

logaritmo do determinante da  $Var(\hat{\beta})$ . Por incorporar este termo, as estimativas produzidas pelo REML são não-viciadas.

As estimativas para os parâmetros dos efeitos fixos costumam não diferir muito se as estimativas de variância forem produzidas pelo REML ou pelo ML. Entretanto, seus erros padrões serão sempre subestimados se os parâmetros de variâncias e covariâncias forem os estimados pelo ML. Assim, o método REML é preferível em relação ao ML (22). Em casos balanceados, a solução obtida através do método REML é idêntica à obtida pelo método da ANOVA (24).

Tanto pelo método ML quanto pelo REML, não há uma equação que especifique as soluções para os parâmetros de variâncias e covariâncias, como existia para os efeitos fixos e aleatórios. Assim, necessita-se de um método iterativo para maximizar as duas funções de verossimilhança, como o algoritmo de Newton-Raphson, disponível no SAS, por exemplo.

Existem outros métodos de estimação de componentes de variância disponíveis, como o estimador quadrático não-viesado de norma mínima (MINQUE) e o estimador quadrático não-viesado de variância mínima (MIVQUE) (24). O primeiro não exige a normalidade do desfecho, suposição presente nos métodos ML e REML, e suas equações não têm que ser resolvidas iterativamente. Se a suposição de normalidade é feita, a solução MINQUE passa a ter a propriedade de variância mínima, sendo chamada então de MIVQUE.

A partir das expressões dos estimadores e das suas variâncias, pode-se fazer inferências tais como intervalos de confiança e teste de hipóteses. Mais detalhes sobre este tópico são dados em Brown e Prescott (22) e na documentação do software SAS (23;25).

### **2.2.3 Modelos para dados longitudinais**

Dados longitudinais são observações feitas seqüencialmente ao longo do tempo nos mesmos sujeitos, indivíduos ou objetos, ou seja, nas mesmas unidades

experimentais. O modelo misto tem características bastante úteis para a análise de dados longitudinais, pois não tem restrições caso os dados sejam não balanceados (medidas obtidas em tempos diferentes nas unidades experimentais, número diferente de medidas repetidas por indivíduo, número de unidades experimentais diferentes por tratamento são casos de desbalanceamento) ou tenham observações faltantes e permite que se modelem correlações entre as observações feitas em uma mesma unidade experimental, o que é bastante coerente.

A aplicação dos modelos mistos para a análise de dados longitudinais foi primeiramente apresentada por Laird e Ware (26), que especificaram o seguinte modelo:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad (2.25)$$

onde

$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ij}, \dots, y_{in_i})$  é o vetor das  $n_i$  observações feitas no sujeito  $i$ ,  $i = 1, \dots, n$ ;

$\mathbf{X}_i$  é uma matriz  $n_i \times p$  conhecida;

$\boldsymbol{\beta}$  é o vetor dos efeitos fixos, de dimensão  $p \times 1$ ;

$\mathbf{Z}_i$  é uma matriz  $n_i \times q$  conhecida;

$\mathbf{u}_i$  é o vetor dos efeitos aleatórios específicos do sujeito  $i$ , de dimensão  $q \times 1$ ;

$\boldsymbol{\varepsilon}_i$  é o vetor de erros aleatórios não-observáveis, de dimensão  $n_i \times 1$ .

Para completar a especificação do modelo, supõe-se que  $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D})$  e  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$ . A matriz  $\mathbf{R}_i$  depende do índice  $i$  apenas em sua dimensão (19;20). Isto significa que, por exemplo, a covariância entre os tempos 1 e 2 é a mesma para todos os indivíduos.

O modelo em (2.25) pode ser especificado da forma de um modelo hierárquico e a hierarquia ajuda a entender as duas fontes de variabilidade existentes no problema envolvendo dados longitudinais. Isto pode ser visto através das expressões a seguir:

$$\text{Estágio 1: } \mathbf{y}_i | \mathbf{u}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i, \mathbf{R}_i)$$

$$\text{Estágio 2: } \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D}).$$

O estágio 1 modela a variabilidade dentro de sujeitos, ou seja, explica-se o porquê da variável resposta não ser constante para as diversas observações no mesmo sujeito. O estágio 2 modela a variabilidade entre sujeitos, ou seja, explica-se o porquê da variável resposta não assumir os mesmos valores para as medidas obtidas no mesmo tempo em sujeitos diferentes.

Há dois tipos de modelos que podem ser utilizados na análise de dados longitudinais, o *modelo de padrão de covariância* e o *modelo de coeficientes aleatórios*. A escolha do tipo de modelo para a análise depende do fato de os intervalos entre as medidas serem fixos (constantes) ou não. Se as medidas ocorrem em intervalos de tempo fixados, pode-se utilizar o modelo de padrão de covariância. Por outro lado, o modelo de coeficientes aleatórios é mais adequado se as medidas ocorrem em tempos não previamente planejados, ou se há particular interesse na relação entre a variável resposta e o tempo (22).

### 2.2.3.1 Modelos de padrão de covariância

Com este modelo pode-se especificar um padrão de covariância que identifique de que maneira as observações de um mesmo sujeito são correlacionadas entre si. Isto pode ser realizado modelando uma estrutura de covariância para as matrizes  $\mathbf{R}_i$  definidas em (2.25). Por exemplo, a estrutura AR(1) especifica que a correlação entre as observações diminui com o tempo. Uma descrição mais abrangente das estruturas de covariância está disponível na literatura (23). Na abordagem clássica, usualmente adota-se a estrutura VC para a matriz  $\mathbf{D}$ . Neste caso, a matriz de covariância  $\mathbf{R}$  do modelo (2.10) é bloco-diagonal, onde os blocos são constituídos pelos indivíduos. A matriz  $\mathbf{R}$  pode ser escrita como

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{R}_2 & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

onde  $\mathbf{R}_i$  são matrizes de covariâncias associadas a cada indivíduo, cuja dimensão é o número de observações de cada sujeito. As matrizes  $\mathbf{0}$  são compostas de zeros e denotam que as observações entre diferentes sujeitos são não correlacionadas.

Há duas abordagens para escolher a estrutura de covariância mais adequada aos dados. A primeira é comparar os modelos através de medidas de qualidade do ajuste, como o Critério de Informação de Akaike (AIC) ou o Critério Bayesiano de Schwarz (BIC), enquanto que a segunda, usualmente mais recomendada, é utilizar o teste da razão de verossimilhanças (22).

### 2.2.3.2 Modelo de coeficientes aleatórios

Neste caso, o modelo descreve aritmeticamente a relação entre uma mensuração e o tempo. As aplicações mais comuns são aquelas em que se assume uma relação linear entre o desfecho (variável resposta) e o tempo. A principal questão de interesse usualmente é avaliar se a taxa de mudança do desfecho é diferente entre os grupos (“tratamentos”).

Assim, um dos efeitos fixos representa a taxa média de mudança (coeficiente angular médio) do desfecho no tempo (isto é, efeito de tempo). A inclusão no modelo da interação tempo\*tratamento, permite verificar em que grau os grupos diferem na taxa média de mudança. Outro efeito fixo que pode ser estimado representa os interceptos médios para cada tratamento (isto é, efeito de tratamento). Em adição aos efeitos fixos que representam os interceptos e coeficientes angulares médios para cada tratamento, o modelo de coeficientes aleatórios permite que os interceptos e coeficientes angulares variem aleatoriamente entre indivíduos, ou seja, uma reta separada é ajustada para cada sujeito. Assim, estima-se o efeito de indivíduo como aleatório, bem como a interação tempo\*indivíduo.

O modelo pode ser especificado com uma estrutura hierárquica, em dois estágios:

$$\text{Estágio 1: } y_{ijk} \mid \alpha_{suj.i}, \beta_{suj.i} \stackrel{iid}{\sim} N(\alpha_{suj.i} + \beta_{suj.i} tempo_j, \sigma_\varepsilon^2) \quad (2.26)$$

$$\text{Estágio 2: } \begin{cases} \alpha_{suj.i} \stackrel{iid}{\sim} N(\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}, \sigma_\alpha^2) \\ \beta_{suj.i} \stackrel{iid}{\sim} N(\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}, \sigma_\beta^2) \end{cases} \quad (2.27)$$

onde  $i = 1, \dots, n$ , onde  $n$  é o número de unidades experimentais (sujeitos);

$j = 1, \dots, n_i$ , onde  $n_i$  é o número observações feitas no sujeito  $i$ ;

$k = 1, \dots, t$ , onde  $t$  é o número de tratamentos;

$t_l, l = 1, \dots, (t-1)$  são variáveis indicadoras dos tratamentos, tal que

$$t_l = \begin{cases} 1, \text{ se } k = l \\ 0, \text{ caso contrário} \end{cases}.$$

É importante notar nas equações (2.26) e (2.27) que a covariável tratamento foi incluída no segundo estágio do modelo. Outras covariáveis tais como idade, peso, etc., também podem ser incluídas no estágio 2 para explicar a variabilidade entre sujeitos.

Este mesmo modelo hierárquico pode ser escrito da forma de um modelo misto apresentada em (2.25), onde  $p = 2t$ ,  $q = 2$  e:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1k} \\ y_{i2k} \\ \vdots \\ y_{in_i k} \end{bmatrix}; \mathbf{u}_i = \begin{bmatrix} \alpha_{suj.i} - (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}) \\ \beta_{suj.i} - (\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}) \end{bmatrix}; \mathbf{Z}_i = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n_i \end{bmatrix} \forall i;$$

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha_0 \\ \beta_0 \\ \vdots \\ \alpha_{(t-1)} \\ \beta_{(t-1)} \end{bmatrix}; \mathbf{X}_i = [\mathbf{Z}_i \mid t_1 \mathbf{Z}_i \mid \dots \mid t_{(t-1)} \mathbf{Z}_i]; \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1k} \\ \varepsilon_{i2k} \\ \vdots \\ \varepsilon_{in_i k} \end{bmatrix}; \mathbf{D} = \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}; \mathbf{R}_i = \sigma_\varepsilon^2 \mathbf{I}_{n_i}.$$

O modelo anterior considera que o intercepto e o coeficiente angular da reta ajustada para cada sujeito são independentes. Entretanto, para se considerar que as observações de um mesmo indivíduo são correlacionadas, pode-se supor que:

$$\mathbf{D} = \begin{bmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_{\beta}^2 \end{bmatrix}$$

ou seja, uma estrutura UN para a matriz  $\mathbf{D}$ .

Cabe ressaltar que o modelo de coeficientes aleatórios permite relações polinomiais (quadrática, cúbica, etc.) do desfecho com o tempo.

O modelo de coeficientes aleatórios pode ser aplicado facilmente quer seja em situações em que as medidas repetidas são observadas em ocasiões no tempo fixadas, quer não. Além disso, sua característica de modelar a relação do desfecho com o tempo é bastante interessante na prática. Por estes motivos, decidiu-se utilizar este modelo neste trabalho.

### 2.3 ABORDAGEM BAYESIANA PARA MODELOS MISTOS NORMAIS

Como foi mencionado, a principal diferença entre a inferência clássica e a inferência Bayesiana é que a primeira supõe que os parâmetros desconhecidos do modelo analisado são constantes, enquanto a segunda os considera quantidades aleatórias. A abordagem Bayesiana combina a informação a priori com a informação proveniente dos dados amostrais através do Teorema de Bayes, dando origem à distribuição a posteriori. Esta distribuição mostra o comportamento dos parâmetros após os dados amostrais serem observados e a partir dela são feitas as inferências sobre os parâmetros e suas devidas interpretações.

Qualquer conjunto de dados pode ser analisado utilizando-se técnicas Bayesianas. Entretanto, alguns problemas, como modelos com muitos parâmetros, exigem cálculos impossíveis de serem executados. Este fato foi um empecilho para a utilização da inferência Bayesiana por muito tempo. Somente com o avanço dos métodos computacionais, especialmente dos métodos MCMC introduzidos na seção

2.1, os métodos Bayesianos puderam ser aplicados a muitos problemas estatísticos, como o caso dos modelos mistos, tratados neste trabalho.

A seção 2.3.1 apresenta brevemente os principais conceitos da inferência Bayesiana, os quais são aplicados, na seção 2.3.2, no contexto de modelos mistos normais. Na seção 2.3.3, as distribuições condicionais completas para o modelo de coeficientes aleatórios são apresentadas.

### 2.3.1 Conceitos básicos de inferência Bayesiana

A inferência Bayesiana está fundamentada na distribuição a posteriori dos parâmetros desconhecidos. Para obter esta distribuição, dois fatores são necessários: a distribuição a priori, denotada por  $\pi(\boldsymbol{\theta})$ , e a função de verossimilhança, denotada por  $L(\boldsymbol{\theta} | \mathbf{y})$ . A distribuição a posteriori resulta de uma aplicação direta do Teorema de Bayes, apresentando a seguinte expressão:

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})}{f(\mathbf{y})},$$

onde  $\boldsymbol{\theta}$  é o vetor de parâmetros desconhecidos,  $\mathbf{y}$  é o vetor de dados observados, e  $f(\mathbf{y}) = \int \cdots \int_{\Theta} \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})d\boldsymbol{\theta}$  é a distribuição marginal de  $\mathbf{Y}$  aplicada no ponto  $\mathbf{y}$ , que é um valor constante em relação a  $\boldsymbol{\theta}$ . Antes de a amostra ser observada,  $f_{\mathbf{Y}}(\cdot)$  é uma função densidade de probabilidade para  $\mathbf{Y}$  chamada de Preditiva a Priori. Seu cálculo envolve, em geral, integração múltipla (no espaço paramétrico) e, por este motivo, muitas vezes não é analiticamente derivável. Note que a denominação distribuição a posteriori vem do fato de que esta é a densidade de probabilidade de  $\boldsymbol{\theta}$  condicional aos dados observados, ou seja, após a realização da amostra.

Um recurso muito utilizado na inferência Bayesiana para determinar a distribuição a posteriori é o da proporcionalidade. Já que  $f(\mathbf{y})$  não depende de  $\boldsymbol{\theta}$ , pode-se estabelecer que

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y}). \quad (2.28)$$

Se através da relação definida em (2.28) for possível identificar o núcleo de alguma distribuição conhecida, sabe-se, então, que a posteriori segue esta distribuição sem necessitar realizar o cálculo de  $f(\mathbf{y})$ . Chama-se núcleo de uma distribuição de probabilidade para  $\boldsymbol{\theta}$  a expressão resultante da retirada de todas as constantes em relação a  $\boldsymbol{\theta}$  da densidade  $f(\boldsymbol{\theta})$ .

Quando a distribuição preditiva a priori não é analiticamente resolvível e o caminho da proporcionalidade não conduz a um núcleo de alguma distribuição conhecida, a posteriori não é analiticamente derivável. Para estes casos, os métodos MCMC são alternativas computacionais que permitem aproximar a distribuição a posteriori de interesse.

A distribuição a priori deve ser escolhida pelo pesquisador, especificando uma forma para esta distribuição e valores para seus parâmetros. Os parâmetros da distribuição a priori são chamados de hiperparâmetros, para não confundir com os parâmetros desconhecidos do modelo. A priori é determinada utilizando todo conhecimento disponível sobre o parâmetro que se está estudando antes de se obter qualquer amostra. Este conhecimento pode ser obtido a partir da opinião de especialistas ou distribuições a posteriori de estudos anteriores, dentre outras fontes. Se não há absolutamente qualquer conhecimento sobre o parâmetro, pode-se atribuir uma priori não-informativa, e assim as estimativas obtidas a partir da distribuição a posteriori serão baseadas apenas na função de verossimilhança. A distribuição a priori não-informativa depende do parâmetro estudado e muitas vezes pode não caracterizar uma distribuição de probabilidade. Isto porque a integral em todos os valores possíveis do parâmetro desconhecido pode não ser igual a 1. Estas prioris são denominadas prioris impróprias. Um exemplo de priori imprópria seria  $\pi(\theta) \propto c$ , onde  $c$  é uma constante. Um cuidado que se deve ter ao especificar priori imprópria é o de averiguar se a respectiva posteriori é realmente uma distribuição de probabilidade. No caso de usar métodos de simulação para estimar os parâmetros é comum não utilizar prioris impróprias, visto que é complicado verificar se a

distribuição a posteriori estimada realmente caracteriza uma distribuição de probabilidade.

Um conceito muito utilizado ao especificar distribuições a priori é o de prioris conjugadas. São prioris que têm a propriedade de se “conjugarem” com a função de verossimilhança, de tal forma que a distribuição a posteriori será da mesma família da distribuição a priori. O uso das prioris conjugadas traz vantagens aos procedimentos de inferência, pois facilita os cálculos analíticos. Prioris conjugadas são sempre prioris próprias. Dependendo dos valores especificados para os hiperparâmetros, uma priori conjugada pode ser não informativa. Por exemplo, uma distribuição normal pode caracterizar uma priori não-informativa se o valor estabelecido para a variância for muito grande.

A distribuição preditiva a priori, definida anteriormente, fornece a distribuição esperada para  $\mathbf{y}$ , já que  $f(\mathbf{y}) = E[L(\boldsymbol{\theta} | \mathbf{y})]$ . Um conceito similar pode ser aplicado para a predição de uma futura observação  $\mathbf{x}$  após  $\mathbf{y}$  ter sido observado. Esta predição deve ser baseada na distribuição de  $\mathbf{x}$  dado  $\mathbf{y}$ , ou seja, na distribuição preditiva a posteriori, definida por

$$\pi(\mathbf{x} | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) L(\boldsymbol{\theta} | \mathbf{x}) d\boldsymbol{\theta}.$$

Nota-se que, analogamente à distribuição preditiva a priori, a preditiva a posteriori fornece a distribuição esperada para  $\mathbf{x} | \mathbf{y}$ .

Com a distribuição a posteriori determinada, podem-se obter estimativas pontuais para cada componente de  $\boldsymbol{\theta}$ , digamos  $\theta_i$ , através do cálculo da esperança, moda ou mediana deste parâmetro, segundo sua distribuição a posteriori marginal definida por

$$\pi(\theta_i | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-i}, \quad (2.29)$$

onde  $\boldsymbol{\theta}_{-i}$  é o vetor  $\boldsymbol{\theta}$  excluindo-se  $\theta_i$ .

O estimador pontual mais utilizado, quando temos uma expressão fechada para a posteriori, é a moda, utilizando conceito análogo ao da maximização de funções de verossimilhança, em que se estima um parâmetro pelo seu valor mais verossímil.

Uma das vantagens da inferência Bayesiana em relação aos métodos clássicos é que as inferências para cada elemento do vetor  $\theta$  são realizadas sem a necessidade de fazer suposições para os demais elementos de  $\theta$ .

Pode-se obter também a estimativa da variância do parâmetro e construir intervalos de credibilidade, calculados diretamente da distribuição a posteriori a partir dos quantis de interesse. Um tipo especial de intervalo de credibilidade, conhecido como HDR (“Highest Density Region”), é construído de tal forma que a densidade de probabilidade dos pontos dentro do intervalo é maior que a de todos os pontos fora dele e a probabilidade do verdadeiro valor do parâmetro estar entre os limites definidos pelo intervalo é igual a  $1-\alpha$ , onde  $0 < \alpha < 1$  é escolhido arbitrariamente. Quando se está estimando a distribuição a posteriori por simulação, não é possível calcular o intervalo HDR. Neste caso utiliza-se intervalos de credibilidade central, calculando os quantis  $\alpha/2$  e  $1-\alpha/2$  da distribuição a posteriori.

Note que, ao utilizar os intervalos de credibilidade, não é necessário obter a distribuição de estimadores como na inferência clássica, que muitas vezes recorre a aproximações destas distribuições. Assim sendo, a inferência Bayesiana possui certas vantagens em relação à clássica, mesmo quando não se tem informação a priori sobre o parâmetro considerado, já que as estimativas pontuais e por intervalo são exatas, quando se tem uma expressão fechada para a distribuição a posteriori, e, também, não necessitam de estimativas de erro padrão e cálculo de distribuições de estimadores.

Muitas vezes, vários modelos são ajustados para tentar explicar da melhor maneira um desfecho. Na análise clássica, existem algumas medidas para se comparar diferentes modelos, como os critérios de Akaike (AIC) e Schwarz (BIC).

Recentemente, Spiegelhalter et. al. (27) sugeriram uma medida para comparar modelos Bayesianos – o “Deviance Information Criterion” (DIC). Essa medida já está disponível no programa WinBUGS e é uma generalização do AIC. O AIC e BIC não podem ser aplicados em alguns problemas Bayesianos em que o número de parâmetros é maior que o número de observações, situação impossível de ser resolvida na abordagem clássica.

Esse critério, assim como outros, compara o ajuste do modelo aos dados e a complexidade do modelo. A medida de ajuste do modelo considerada no DIC é a função desvio (“Deviance”), definida como  $D(\theta) = -2\log L(\theta/y)$ . A complexidade do modelo é estimada como o número efetivo de parâmetros  $p_D$

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}),$$

isto é, a média a posteriori da “deviance” menos a “deviance” avaliada na média a posteriori dos parâmetros. O DIC é então definido como

$$DIC = D(\bar{\theta}) + 2p_D = \bar{D} + p_D.$$

O modelo com o menor DIC é aquele que faria melhores previsões em um conjunto de dados replicado com a mesma estrutura do conjunto de dados observado. Entretanto, se a diferença em DIC é menor que 5, e os modelos fazem inferências muito diferentes, então pode ser enganador apenas reportar o modelo com o menor DIC (28).

### 2.3.2 Análise Bayesiana para modelos mistos normais

Para se fazer uma análise estatística através das técnicas Bayesianas para o modelo misto normal especificado em (2.10), primeiramente precisa-se definir a distribuição a priori conjunta para os parâmetros desconhecidos, que são  $\beta$ ,  $\mathbf{D}$  e  $\mathbf{R}$ . Ou seja, precisa-se determinar a priori conjunta  $\pi(\beta, \mathbf{D}, \mathbf{R})$ . Usualmente, supõe-se independência a priori entre estes parâmetros, tal que  $\pi(\beta, \mathbf{D}, \mathbf{R}) = \pi(\beta)\pi(\mathbf{D})\pi(\mathbf{R})$ . Assumindo independência a priori, necessita-se especificar distribuições para os

parâmetros  $\boldsymbol{\beta}$ ,  $\mathbf{D}$  e  $\mathbf{R}$  separadamente. Note que o vetor de efeitos aleatórios  $\mathbf{u}$  já tem sua distribuição de probabilidade especificada pelas suposições do modelo.

Segundo a formulação do modelo misto visto como um modelo hierárquico definida nas equações (2.11) e (2.12), as distribuições a priori  $\pi(\boldsymbol{\beta})$ ,  $\pi(\mathbf{D})$  e  $\pi(\mathbf{R})$  correspondem a um terceiro estágio da hierarquia.

A distribuição a priori conjugada usual para  $\boldsymbol{\beta}$  é

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}^*, \mathbf{H}),$$

onde  $\boldsymbol{\beta}^*$  é um vetor conhecido, de dimensão  $p \times 1$ , e  $\mathbf{H}$  é uma matriz de variâncias e covariâncias conhecida, de dimensão  $p \times p$ . Uma priori não-informativa para  $\boldsymbol{\beta}$  pode ser obtida especificando-se  $\mathbf{H} = \sigma^2 \mathbf{I}_p$ , onde  $\sigma^2 \rightarrow \infty$ .

As distribuições a priori para  $\mathbf{D}$  e  $\mathbf{R}$  dependerão da estrutura de covariância adotada para cada uma delas. Em modelos normais, o mais usual é parametrizar componentes de variância e matrizes de variâncias e covariâncias como componentes e matrizes de precisão, pois isto facilita os cálculos das distribuições a posteriori. Um componente de precisão é o recíproco do componente de variância correspondente, e uma matriz de precisão é a inversa da correspondente matriz de variância.

Se a estrutura de covariância VC é adotada para a distribuição a priori para os componentes e matrizes de precisão, então a distribuição Gama pode ser usada como priori para cada um dos componentes de precisão da matriz, ou seja,  $\sigma_w^{-2} \sim G(a_w, b_w)$ . A priori não-informativa neste caso é obtida quando  $a_w, b_w \rightarrow 0$ . Outra distribuição a priori não-informativa que poderia ser obtida especificando  $\sigma_w \sim U(0, c_w)$ , onde  $c_w \rightarrow \infty$ . Neste caso, a distribuição de probabilidade para  $\sigma_k^{-2}$

tem densidade  $f(\sigma_w^{-2}) = \frac{(\sigma_w^{-2})^{-3/2}}{2c_w}$ ,  $c_w^{-2} \leq \sigma_w^{-2} < +\infty$ . Esta última tem sido

considerada uma opção melhor que a priori Gama principalmente para componentes de variância da matriz  $\mathbf{D}$  (29).

Se uma estrutura AR(1) é usada para a matriz de variâncias e covariâncias, pode-se especificar uma distribuição Gama para o componente de precisão  $\sigma^{-2}$ , e uma priori para o parâmetro  $\rho$  poderia ser  $\rho \sim U(a, b)$ , onde  $a < b$  e  $a \geq -1$  e  $b \leq 1$ . Se  $a = -1$  e  $b = 1$  são especificados, então se obtém uma distribuição a priori não-informativa para  $\rho$ .

No caso da estrutura UN, não é necessário assumir uma priori para cada componente de variância da matriz. As distribuições Wishart e Inversa-Wishart são distribuições de probabilidade para, respectivamente, as matrizes de precisão e de variâncias e covariâncias. Além disso, são prioris conjugadas para, respectivamente, a matriz de precisão e de variâncias e covariâncias de uma distribuição normal multivariada. Segundo Gelman et al. (30), se  $W \sim Wishart(\rho, S)$ , em que  $W$  é uma matriz positiva-definida simétrica  $k \times k$ ,  $\rho$  é o número de graus de liberdade e  $S$  é uma matriz de escala positiva-definida simétrica  $k \times k$ , então a função densidade de probabilidade é dada por

$$f_w(w) = \left( 2^{\rho k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\rho+1-i}{2}\right) \right)^{-1} |S|^{-\rho/2} |w|^{(\rho-k-1)/2} \exp\left(-\frac{1}{2} tr(S^{-1}w)\right).$$

Da mesma forma, se  $W \sim Inv-Wishart(\rho, S^{-1})$ , em que  $W$  é uma matriz positiva-definida simétrica  $k \times k$ ,  $\rho$  é o número de graus de liberdade e  $S$  é uma matriz de escala positiva-definida simétrica  $k \times k$ , então a função densidade de probabilidade é dada por

$$f_w(w) = \left( 2^{\rho k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\rho+1-i}{2}\right) \right)^{-1} |S|^{\rho/2} |w|^{-(\rho+k+1)/2} \exp\left(-\frac{1}{2} tr(Sw^{-1})\right).$$

A esperança da distribuição Wishart é  $\rho S$ , e da Inversa-Wishart é  $(\rho - k - 1)^{-1} S$ . Por isso, nas aplicações Bayesianas, quando se deseja atribuir uma priori Wishart para uma matriz de precisão, geralmente é parametrizada como  $Wishart(\rho, (\rho D^*)^{-1})$ , onde  $D^*$  é uma matriz  $k \times k$ . Assim,  $D^*$  será a média a priori da matriz de variâncias que se deseja estimar, e seus valores podem ser escolhidos conforme as expectativas para os possíveis valores da matriz de variâncias. Uma

priori não-informativa tanto para a distribuição Wishart quanto para a Inversa-Wishart é obtida quando  $\rho = k$ .

Após a especificação das distribuições a priori apropriadas para os parâmetros  $\beta$ ,  $\mathbf{D}$  e  $\mathbf{R}$ , levando em consideração as estruturas de covariância assumidas para as matrizes  $\mathbf{D}$  e  $\mathbf{R}$ , o próximo passo seria a derivação da distribuição a posteriori. Porém, independente da estrutura escolhida, não é possível determinar analiticamente a distribuição a posteriori, ou reconhecer o núcleo de alguma distribuição multivariada conhecida através do caminho da proporcionalidade. O algoritmo Amostrador de Gibbs pode ser aplicado neste caso, já que, para os modelos mistos, é possível derivar as distribuições condicionais completas.

### 2.3.3 Distribuições condicionais completas do modelo de coeficientes aleatórios

As distribuições condicionais completas de todos os parâmetros desconhecidos são necessárias para o Amostrador de Gibbs. Estas distribuições também dependem das distribuições a priori especificadas. A seguir, serão mostradas as distribuições condicionais para o modelo de coeficientes aleatórios, considerando o modelo hierárquico definido em (2.26) e (2.27), ou seja,

$$\text{Estágio 1: } y_{ijk} | \alpha_{suj.i}, \beta_{suj.i} \stackrel{iid}{\sim} N(\alpha_{suj.i} + \beta_{suj.i} tempo_j, \sigma_\epsilon^2)$$

$$\text{Estágio 2: } \begin{cases} \alpha_{suj.i} \stackrel{iid}{\sim} N(\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}, \sigma_\alpha^2) \\ \beta_{suj.i} \stackrel{iid}{\sim} N(\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}, \sigma_\beta^2) \end{cases},$$

através da abordagem Bayesiana determinando prioris para os parâmetros  $\alpha_0, \dots, \alpha_{t-1}, \beta_0, \dots, \beta_{t-1}, \sigma_\epsilon^{-2}, \sigma_\alpha^{-2}$  e  $\sigma_\beta^{-2}$ . Os interceptos e coeficientes angulares de cada sujeito já possuem uma distribuição a priori atribuída, inerente ao modelo. Para os componentes de precisão, pode-se especificar as mesmas distribuições a priori do

caso da estrutura de covariância VC. Para os parâmetros  $\alpha_s$  e  $\beta_s$ ,  $s=0, \dots, t-1$ , pode-se especificar distribuições normais.

Assim, considere as distribuições a priori

$$\alpha_h \sim N(a_h, c_h^2), h = 0, 1, 2;$$

$$\beta_h \sim N(b_h, d_h^2), h = 0, 1, 2;$$

$$\sigma_\varepsilon^{-2} \sim G(e_1, e_2);$$

$$\sigma_\alpha^{-2} \sim f(\sigma_\alpha^{-2}) = \frac{(\sigma_\alpha^{-2})^{-3/2}}{2c_\alpha}, c_\alpha^{-2} \leq \sigma_\alpha^{-2} < +\infty \text{ e}$$

$$\sigma_\beta^{-2} \sim f(\sigma_\beta^{-2}) = \frac{(\sigma_\beta^{-2})^{-3/2}}{2c_\beta}, c_\beta^{-2} \leq \sigma_\beta^{-2} < +\infty.$$

Note que se está particularizando o modelo para o caso de 3 tratamentos, pois é a situação escolhida para ser simulada no artigo. Com algumas manipulações algébricas chega-se às condicionais completas a seguir.

$$\pi(\alpha_0 | \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{c_0^2 \left( \sum_{i=1}^n \alpha_{suj.i} - n_{trat1} \alpha_1 - n_{trat2} \alpha_2 \right) + a_0 \sigma_\alpha^2}{nc_0^2 + \sigma_\alpha^2}; \frac{c_0^2 \sigma_\alpha^2}{nc_0^2 + \sigma_\alpha^2} \right);$$

$$\pi(\alpha_1 | \alpha_0, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{c_1^2 \left( \sum_{i=1}^n t_1 \alpha_{suj.i} - n_{trat1} \alpha_0 \right) + a_1 \sigma_\alpha^2}{n_{trat1} c_1^2 + \sigma_\alpha^2}; \frac{c_1^2 \sigma_\alpha^2}{n_{trat1} c_1^2 + \sigma_\alpha^2} \right);$$

$$\pi(\alpha_2 | \alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{c_2^2 \left( \sum_{i=1}^n t_2 \alpha_{suj.i} - n_{trat2} \alpha_0 \right) + a_2 \sigma_\alpha^2}{n_{trat2} c_2^2 + \sigma_\alpha^2}; \frac{c_2^2 \sigma_\alpha^2}{n_{trat2} c_2^2 + \sigma_\alpha^2} \right);$$

$$\pi(\beta_0 | \alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{d_0^2 \left( \sum_{i=1}^n \beta_{suj.i} - n_{trat1} \beta_1 - n_{trat2} \beta_2 \right) + b_0 \sigma_\beta^2}{n d_0^2 + \sigma_\beta^2}; \frac{d_0^2 \sigma_\beta^2}{n d_0^2 + \sigma_\beta^2} \right);$$

$$\pi(\beta_1 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{d_1^2 \left( \sum_{i=1}^n t_1 \beta_{suj.i} - n_{trat1} \beta_0 \right) + b_1 \sigma_\beta^2}{n_{trat1} d_1^2 + \sigma_\beta^2}; \frac{d_1^2 \sigma_\beta^2}{n_{trat1} d_1^2 + \sigma_\beta^2} \right);$$

$$\pi(\beta_2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{d_2^2 \left( \sum_{i=1}^n t_2 \beta_{suj.i} - n_{trat2} \beta_0 \right) + b_2 \sigma_\beta^2}{n_{trat2} d_2^2 + \sigma_\beta^2}; \frac{d_2^2 \sigma_\beta^2}{n_{trat2} d_2^2 + \sigma_\beta^2} \right);$$

$$\pi(\alpha_{suj.i} | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, i-1, i+1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{\sigma_\alpha^2 \sum_{i=1}^n (y_{ijk} - \beta_{suj.i} tempo_j) + \sigma_\varepsilon^2 (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2)}{g \sigma_\alpha^2 + \sigma_\varepsilon^2}; \frac{\sigma_\alpha^2 \sigma_\varepsilon^2}{g \sigma_\alpha^2 + \sigma_\varepsilon^2} \right);$$

$$\pi(\beta_{suj.i} | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, i-1, i+1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left( \frac{\sigma_\beta^2 \sum_{i=1}^n tempo_j (y_{ijk} - \alpha_{suj.i}) + \sigma_\varepsilon^2 (\beta_0 + \beta_1 t_1 + \beta_2 t_2)}{\sigma_\beta^2 \sum_{j=1}^g tempo_j^2 + \sigma_\varepsilon^2}; \frac{\sigma_\beta^2 \sigma_\varepsilon^2}{\sigma_\beta^2 \sum_{j=1}^g tempo_j^2 + \sigma_\varepsilon^2} \right);$$

$$\pi(\sigma_\alpha^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$IG \left( \frac{n-1}{2}; \frac{\sum_{i=1}^n (\alpha_{suj.i} - (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2))^2}{2} \right);$$

$$\pi(\sigma_\beta^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\varepsilon^2) \sim$$

$$IG \left( \frac{n-1}{2}; \frac{\sum_{i=1}^n (\beta_{suj.i} - (\beta_0 + \beta_1 t_1 + \beta_2 t_2))^2}{2} \right) e$$

$$\pi(\sigma_\varepsilon^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2) \sim$$

$$IG \left( \frac{ng}{2} + e_1; \frac{\sum_{i=1}^n \sum_{j=1}^g (y_{ijk} - (\alpha_{suj.i} + \beta_{suj.i} tempo_j))^2}{2} + e_2 \right).$$

Com estas distribuições condicionais, a aplicação do Amostrador de Gibbs pode ser feita em qualquer programa que gere valores aleatórios a partir das distribuições Normal e Gama, já que para gerar números de uma distribuição Inversa-Gama somente é necessário tomar o recíproco dos valores gerados de uma Gama (30).

É importante destacar que, utilizando o programa WinBUGS, pode-se analisar estes e outros problemas Bayesianos através do Amostrador de Gibbs sem a necessidade de conhecer as expressões das distribuições condicionais completas,

bastando informar as distribuições a priori e a função de verossimilhança. Uma breve descrição deste programa encontra-se no Anexo 4.

O programa R também possui alguns recursos úteis para se utilizar o Amostrador de Gibbs. Dentre eles destaca-se o pacote *MCMCpack*, disponível na Internet, que possui funções como *rwish* que gera números aleatórios da distribuição Wishart, além de funções que aplicam MCMC a modelos de regressão linear, logística, entre outros.

### 3 REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Leotti VB, Ziegelmann PK. Abordagem Bayesiana para Modelos Mistos com o Uso de "Gibbs Sampling" [Monografia para obtenção do título de Bacharel em Estatística.] Departamento de Estatística - UFRGS; 2005.
- (2) Browne WJ. Applying MCMC Methods to Multilevel Models Department of Mathematical Sciences, University of Bath, UK.; 1998.
- (3) Browne WJ, Draper D. A comparison of Bayesian and likelihood methods for fitting multilevel models. *Bayesian Analysis* 2006;1(3):473-514.
- (4) Browne WJ, Draper D. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics* 2000;15(3):391-420.
- (5) Gamerman D. Markov Chain Monte Carlo. London: Chapman & Hall; 1997.
- (6) Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in Practice. London: Chapman & Hall; 1996.
- (7) Ross S. Stochastic Processes. Nova York: John Wiley & Sons, Inc.; 1996.
- (8) Hastings WK. Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 1970;57(1):97-&.
- (9) Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* 1953;21(6):1087-92.
- (10) Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 1984;6(6):721-41.
- (11) Gelfand AE, Smith AFM. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 1990 Jun;85(410):398-409.
- (12) Tierney L. Markov-Chains for Exploring Posterior Distributions. *Annals of Statistics* 1994 Dec;22(4):1701-28.
- (13) Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* 1996 Jun;91(434):883-904.
- (14) Brooks SP, Roberts GO. Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing* 1998 Dec;8(4):319-35.

- (15) Best N, Cowles M, Vines S. CODA: Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Version 0.3. Cambridge, Inglaterra.: Biostatistics Unit-MRC; 1995.
- (16) Spiegelhalter DJ, Thomas A, Best N, Gilks WR. Bayesian Inference Using Gibbs Sampling: Version 0.5. Cambridge, Inglaterra: Biostatistics Unit-MRC; 1995.
- (17) Casella G, George EI. Explaining the Gibbs Sampler. *American Statistician* 1992 Aug;46(3):167-74.
- (18) Henderson C. Applications of Linear Models in Animal Breeding. Ontario: University of Guelph; 1984.
- (19) Davidian M, Giltinan DM. Nonlinear Models for Repeated Measurement Data. San Francisco: Chapman & Hall; 1995.
- (20) Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer Verlag; 2000.
- (21) Bryk AS, Raudenbush SW. Hierarchical Linear Models. Sage; 1992
- (22) Brown H, Prescott R. Applied Mixed Models in Medicine. 2nd ed. Chichester: John Wiley & Sons, Inc.; 2006.
- (23) SAS Institute Inc. SAS Online Doc v.9. Cary, EUA: SAS Institute Inc.; 2002.
- (24) Searle S, Casella G, McCulloch C. Variance Components. New York: John Wiley & Sons, Inc.; 1992.
- (25) Littell RC, et.al. SAS for mixed models. 2nd ed. Cary, N.C: SAS Institute, Inc; 2006.
- (26) Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982 Dec;38(4):963-74.
- (27) Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2002;64:583-616.
- (28) The BUGS Project. Frequently Asked Questions about DIC. Web Page 2007 November 9 Available from: URL: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml>
- (29) Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006;1(3):515-34.
- (30) Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. Second ed. Boca Raton: Chapman & Hall; 2004.

## **4 OBJETIVOS**

Realizar um estudo de simulação de dados longitudinais para avaliar vício e precisão das estimativas obtidas através das análises clássica e Bayesiana para o modelo de coeficientes aleatórios.

## 5 ARTIGO

### **Comparação via Simulação dos Estimadores Clássicos e Bayesianos no Modelo de Coeficientes Aleatórios para Dados Longitudinais**

*A comparison by Simulation of Bayesian and Classical Estimators in the Random  
Coefficients Model for Longitudinal Data*

Vanessa Bielefeldt Leotti <sup>1,2</sup>

Patrícia Klarmann Ziegelmann <sup>1</sup>

João Riboldi <sup>1,2</sup>

Álvaro Vigo <sup>1,2</sup>

<sup>1</sup> Departamento de Estatística, Instituto de Matemática, UFRGS.

<sup>2</sup> Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina, UFRGS

**A ser enviado ao Computational Statistics and Data Analysis**

## Resumo

Frequentemente em pesquisas médicas ou epidemiológicas, múltiplas medidas de um mesmo sujeito são tomadas ao longo do tempo, caracterizando um estudo com dados longitudinais. Nos últimos anos, a técnica estatística que tem sido mais utilizada para a análise desses estudos é o modelo misto, pois este permite que a provável correlação das observações de um mesmo indivíduo seja modelada, e é flexível frente às situações de desbalanceamento e dados faltantes. Nesta classe, o modelo de coeficientes aleatórios permite que a relação da variável resposta com o tempo seja descrita através de uma função matemática. Inferências para os parâmetros do modelo podem ser conduzidas sob enfoque da estatística clássica ou Bayesiana. As duas abordagens podem conduzir a resultados diferentes, deixando o pesquisador em dúvida sobre a escolha do método. Há poucos estudos na literatura comparando as abordagens clássica e Bayesiana para a análise de dados longitudinais via modelos mistos. O objetivo deste trabalho foi proceder a um estudo de simulação, para comparar estas abordagens em termos de vício e precisão. Foram comparados os estimadores Bayesianos média, moda (estimada não-parametricamente) e mediana a posteriori, obtidos através do Amostrador de Gibbs, e o estimador clássico obtido pelo método REML. O modelo simulado envolve efeito de tempo e uma covariável denominada tratamento, e assume a estrutura Componentes de Variância para as duas matrizes de variância do modelo. Diferentes configurações de tamanho de amostra e desbalanceamento foram adotadas, para avaliar o desempenho dos métodos frente a essas situações. As maiores diferenças encontradas foram em relação a alguns componentes de variância, sendo que pelo menos um dos estimadores Bayesianos apresentou erro quadrático médio menor do que o estimador clássico em todas as configurações. Não foi possível identificar um único estimador Bayesiano como sendo o melhor para todos os casos estudados. A moda mostrou-se um estimador com boas propriedades em algumas situações e por isso sugere-se que o mesmo seja implementado no software WinBUGS. Um fato interessante é que todos os estimadores tiveram melhor desempenho na situação onde os intervalos de tempo entre medições não eram constantes.

## Abstract

Frequently in medical or epidemiologic research, multiple measures of the same subject are observed over the time. This characterizes a study with longitudinal data. In the last years, the statistical technique that has been most used in the analysis of these studies is the mixed model, because it permits to model the probable correlation between the observations of the same individual, and it is flexible in situations of unbalanceament and missing data. A mixed model that has been used in these cases is the random coefficients model. It allows the relationship between the response variable and the time to be described through a mathematical function. Inferences for the model parameters can be conducted with the frequentist approach or the Bayesian one. The two approaches may produce different results, and because of that the researcher may be in doubt of what method to use. In the literature, there are few studies that compare the classical and the Bayesian approach in the analysis of longitudinal data using mixed models with an analytical way. So, the objective of this work is to proceed a simulation study to compare the Bayesian and classical inferences in terms of bias and precision. The Bayesian estimators posterior mean, (non-parametric) mode and median, obtained using the Gibbs Sampler, and the classical estimator obtained by the REML method were compared. The simulated model involves the effects of time and a covariable named treatment, and that assumes a Variance Components structure for the two covariance matrices of the mixed model. Different configurations of sample size and unbalanced data were adopted, to see the performance of the methods with these situations. The main differences founded were in the estimation of some variance components. For these parameters, one of the Bayesian estimators had lower mean square error than the classical in all of the configurations. It was not possible to identify one best Bayesian estimator in all the cases studied. The mode seems an estimator with good properties in some situations and because of that it should be implemented in the WinBUGS software. One interesting fact is that all the estimators had better performance in the situation where the time intervals between observations were not fixed.

**Palavras chaves**

Dados longitudinais, modelos mistos, modelo de coeficientes aleatórios, Amostrador de Gibbs, MCMC, simulação.

**Número de Palavras: 4195****Introdução**

Na pesquisa médica e epidemiológica, é cada vez mais comum encontrar estudos que utilizam dados longitudinais. Esse tipo de dados ocorre quando múltiplas medidas de um desfecho são tomadas em uma mesma unidade experimental (em geral, indivíduos) em várias ocasiões no tempo. Esses estudos permitem ao pesquisador descrever o comportamento da variável resposta ao longo do tempo, verificar se existe tendência de crescimento ou decrescimento, comparar a eficácia de diferentes intervenções ao longo do tempo, bem como comparar diferenças entre grupos (1).

Nos últimos anos, a técnica estatística que tem sido mais utilizada para a análise de dados longitudinais é o modelo misto. Esse modelo compreende efeitos fixos e aleatórios em sua formulação e assume distribuição multivariada tanto para os efeitos aleatórios como para os erros. Esta última característica permite que dados que não são independentes sejam analisados. Outra boa propriedade do modelo misto é sua grande flexibilidade frente às situações com dados faltantes e desbalanceamento. Em dados longitudinais, três tipos de desbalanceamento podem ocorrer: número diferente de indivíduos em cada grupo de comparação, número diferente de medidas repetidas tomadas em cada indivíduo e observações realizadas em ocasiões no tempo diferentes.

Há dois casos especiais do modelo misto que são úteis para se analisar dados longitudinais. O primeiro é o modelo de padrão de covariância, onde se assume um padrão para a matriz de variâncias e covariâncias do erro aleatório. Este modelo geralmente é utilizado quando as medidas repetidas são tomadas em intervalos de tempo fixados. O segundo é o modelo de coeficientes aleatórios, onde se ajusta uma

função matemática para descrever o comportamento da variável resposta ao longo do tempo. Este modelo pode ser aplicado facilmente quando os intervalos de tempo entre as medidas não são fixos, e pode-se introduzir a correlação das medidas de um mesmo indivíduo permitindo que os efeitos aleatórios tenham uma distribuição multivariada (1). Pelo fato do modelo de coeficientes aleatórios poder ser utilizado tanto em situações de balanceamento em relação às ocasiões no tempo em que as medidas são observadas, quanto nas situações de desbalanceamento, e considerando que a sua característica de modelar a relação do desfecho com o tempo é bastante interessante na prática, decidiu-se focar este modelo neste trabalho.

A classe dos modelos mistos pode ser analisada por duas abordagens estatísticas: a clássica (1) e a Bayesiana (2;3). De uma maneira geral, a principal diferença entre inferência Bayesiana e inferência clássica é que a primeira considera os parâmetros desconhecidos como quantidades aleatórias, enquanto a segunda os considera constantes. Segundo a abordagem Bayesiana, antes de qualquer dado ser coletado deve-se assumir uma distribuição de probabilidade para os parâmetros desconhecidos do modelo, chamada de distribuição a priori, a qual deve representar toda a informação existente sobre estes parâmetros. Após os dados serem observados, a informação disponível é combinada, através do teorema de Bayes, com a informação proveniente dos dados, que é representada pela função de verossimilhança associada à amostra observada. Esta combinação resulta na distribuição a posteriori dos parâmetros desconhecidos que proverá toda a informação sobre estes após os dados terem sido observados. Assim, toda a inferência desejada será retirada da posteriori. Quando não há informação disponível sobre alguns dos parâmetros, pode-se utilizar prioris não-informativas para os mesmos, e então os resultados da análise Bayesiana serão baseados apenas na função de verossimilhança, e assim serão comparáveis aos resultados da análise clássica.

O modelo misto não possui distribuição a priori analiticamente derivável. Por isto, deve-se utilizar um dos métodos MCMC (acrônimo de “Markov Chain Monte Carlo”) para se simular a distribuição a posteriori. Neste trabalho, utilizou-se o Amostrador de Gibbs (“Gibbs Sampler”), já que as distribuições condicionais completas necessárias para a implementação do algoritmo são conhecidas.

Na análise de dados longitudinais, as abordagens Bayesiana e clássica podem levar a resultados diferentes, e assim, o pesquisador pode ficar em dúvida sobre qual método utilizar. Na literatura, há poucos estudos que comparam as duas abordagens para dados longitudinais sob um enfoque analítico. A maioria das discussões existentes aponta vantagens e desvantagens de cada um dos métodos apenas sob o ponto de vista conceitual. Alguns dos argumentos em favor da abordagem Bayesiana são: a) na inferência clássica para os modelos mistos, deve-se assumir que as variâncias populacionais são conhecidas para poder obter uma expressão para a variância dos estimadores, uma suposição que é raramente atendida. A consequência desse fato é um vício nos erros padrões dos efeitos fixos e aleatórios (1). Isto já não ocorre na análise Bayesiana, pois não é necessário supor as variâncias conhecidas para realizar a análise; b) na análise clássica para modelos mistos, os parâmetros são estimados por algoritmos iterativos que podem ter problemas com amostras pequenas, muitos dados faltantes, etc. A inferência Bayesiana geralmente se comporta melhor frente a essas situações; e, c) no caso dos parâmetros de variâncias do modelo misto, às vezes a análise clássica produz estimativas negativas de variâncias (método da ANOVA) ou, então, o método iterativo trunca e não fornece nenhuma estimativa (métodos ML e REML). Na inferência Bayesiana, geralmente esses problemas não ocorrem. Uma desvantagem da abordagem Bayesiana é a sua implementação computacional. O software gratuito WinBUGS já está com uma interface mais amigável do que sua primeira versão (4), mas ainda exige algum entendimento de programação do usuário.

Browne (5) e Browne e Draper (6;7) procederam a estudos de simulação para alguns casos especiais do modelo misto e também do modelo misto generalizado, mas não trabalharam no contexto de dados longitudinais. Um dos modelos estudados (5;6) foi um modelo de ANOVA com um fator aleatório, que é mais simples que os modelos para dados longitudinais. Outro modelo estudado (5;7) foi um modelo multinível (hierárquico) de 2 níveis, chamado pelos autores de modelo de regressão de coeficientes aleatórios. As simulações foram baseadas em um problema real em que se tinham escolas e alunos dentro das escolas, e o objetivo foi relacionar as notas dos alunos em um teste de matemática em dois determinados anos. Assim, o modelo ajustava uma reta para descrever essa relação, mas cada escola tinha sua particular

reta, pois se considerou escola como um efeito aleatório. Esse modelo é idêntico ao modelo de coeficientes aleatórios para dados longitudinais, pois as escolas podem ser comparadas com os indivíduos, e os alunos dentro das escolas podem ser comparados com as observações para um mesmo indivíduo. Entretanto, esse modelo não tem covariáveis, como usualmente ocorre em problemas de dados longitudinais.

No contexto de dados longitudinais, um estudo de simulação (8), foi realizado para comparar os estimadores Bayesianos média, moda (estimada não-parametricamente) e mediana a posteriori, mas não foram comparados com os estimadores clássicos. O modelo simulado considerou a estrutura Componentes de Variância para ambas as matrizes de covariância dos efeitos aleatórios e do erro. Além disso, considerou a situação de simular dados com uma estrutura Não-estruturada para a matriz de covariância do erro, mas analisar os mesmos supondo uma estrutura Componentes de Variância (isto é, erros independentes). No estudo conclui-se que supor erros independentes quando na verdade os erros estão correlacionados e as variâncias são diferentes, os estimadores dos efeitos fixos e dos componentes de variância se tornam viesados. Conclui também que a moda estimada não-parametricamente é um estimador com boas propriedades no caso do modelo estudado.

Neste sentido, o objetivo deste trabalho foi realizar um estudo de simulação para dados longitudinais para comparar as propriedades dos estimadores clássicos e Bayesianos em termos de vício e precisão.

## Métodos

Para o estudo de simulação, considerou-se um modelo que envolve múltiplas medidas de um mesmo indivíduo (efeito de tempo) e uma única covariável que define grupos de indivíduos (efeito de tratamento). O modelo considerado foi um modelo de coeficientes aleatórios, formulado como

$$y_{ijk} | \alpha_{suj.i}, \beta_{suj.i} \sim N(\alpha_{suj.i} + \beta_{suj.i} tempo_j, \sigma_\epsilon^2)$$

$$\alpha_{suj.i} \stackrel{iid}{\sim} N(\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2, \sigma_\alpha^2) \text{ e } \beta_{suj.i} \stackrel{iid}{\sim} N(\beta_0 + \beta_1 t_1 + \beta_2 t_2, \sigma_\beta^2),$$

onde  $i = 1, \dots, n$  e  $n$  é o número de unidades experimentais (sujeitos);  $j = 1, \dots, n_i$  sendo que  $n_i$  é o número de observações feitas no sujeito  $i$ ;  $k = 1, \dots, t$  onde  $t$  é o número de tratamentos ( $t=3$ );  $t_l, l = 1, 2$ , são variáveis indicadoras dos tratamentos, tal que o terceiro tratamento é a categoria de referência.

O modelo acima permite que cada indivíduo tenha uma particular reta, e que essa reta seja influenciada pelo tratamento para o qual foi alocado. Cada tratamento também tem a sua particular reta, com intercepto  $\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2$  e coeficiente angular  $\beta_0 + \beta_1 t_1 + \beta_2 t_2$ . Este modelo está assumindo correlação nula entre as observações do mesmo sujeito, já que se considerou a estrutura de covariância Componentes de Variância para ambas as matrizes de covariância dos efeitos aleatórios e do erro, embora isto não seja comum na prática.

Foram adotadas três configurações para o número de ocasiões no tempo e de indivíduos dentro de cada tratamento, representando situações de amostra pequena e desbalanceamento. Nas três situações, o número de tratamentos,  $t$ , era igual a 3.

- Configuração 1:  $n = 27$  sujeitos, sendo 8 no tratamento 1, 10 no tratamento 2 e 9 no tratamento 3;  $n_i = 5, \forall i$  e  $tempo = (1 \ 2 \ 3 \ 4 \ 5)^t \forall i$ . Esta configuração representa uma situação desbalanceada em relação ao número de indivíduos em cada tratamento, mas balanceada em relação às ocasiões no tempo, e com tamanho de amostra pequeno.

- Configuração 2:  $n = 27$  sujeitos, sendo 8 no tratamento 1, 10 no tratamento 2 e 9 no tratamento 3;  $n_i = 5, \forall i$  mas o vetor  $tempo$  é constituído de 5 valores sorteados de uma Uniforme Discreta com valores entre 1 a 15, inclusive. Esta configuração representa um problema desbalanceado em relação ao número de indivíduos em cada tratamento e às ocasiões no tempo, com tamanho de amostra pequeno.

- Configuração 3:  $n = 60$  sujeitos, sendo 20 em cada tratamento;  $n_i = 5, \forall i$  e  $tempo = (1 \ 2 \ 3 \ 4 \ 5)^t \forall i$ , representando uma situação balanceada tanto em

relação ao número de pessoas em cada tratamento e às ocasiões no tempo, com tamanho de amostra razoavelmente grande.

Essas situações foram escolhidas por intuitivamente apresentar relevância prática no contexto das pesquisas médicas. Em geral, planejam-se inicialmente estudos com igual número de indivíduos em cada tratamento, mas ao decorrer do mesmo, ocorrem perdas que provocam um pequeno desbalanceamento em relação à distribuição das unidades nos tratamentos. Usualmente, é mais fácil controlar o número de medidas repetidas que se vai tomar do que exatamente a ocasião no tempo em que a medição será realizada. Ou seja, entre as 3 configurações, a configuração 2 é mais provável de ocorrer na prática. Além disso, por razões éticas os estudos com medidas repetidas em ensaios clínicos usualmente não consideram amostras de tamanho muito grande, por isso o maior tamanho de amostra considerado aqui no estudo foi 60.

Nas três situações, os valores determinados para os parâmetros do modelo foram:  $\alpha_0 = 1$ ;  $\alpha_1 = 5$ ;  $\alpha_2 = 3$ ;  $\beta_0 = 2$ ;  $\beta_1 = -4$ ;  $\beta_2 = -6$ ;  $\sigma_\alpha^2 = 1$ ;  $\sigma_\beta^2 = 4$ ;  $\sigma_\varepsilon^2 = 9$ . Ou seja, especificou-se  $\alpha_{trat.1} = 6$ ,  $\alpha_{trat.2} = 4$ ,  $\alpha_{trat.3} = 1$ ,  $\beta_{trat.1} = -2$ ,  $\beta_{trat.2} = -4$  e  $\beta_{trat.3} = 2$ . A partir desses valores, 1000 conjuntos de dados foram gerados aleatoriamente conforme a estrutura do modelo. Para cada conjunto, as análises clássicas e Bayesianas foram procedidas. A análise clássica utilizou o método REML de estimação de componentes de variância. Não se considerou o método ML, pois já está claro na literatura (1;5-7) que este método é inferior ao REML, produzindo estimativas viciadas dos componentes de variância.

Para a análise Bayesiana, as seguintes prioris não-informativas foram consideradas:  $\alpha_h \sim N(0;10000)$ ,  $h = 0,1,2$ ;  $\beta_h \sim N(0;10000)$ ,  $h = 0,1,2$ ;  $\sigma_\varepsilon^{-2} \sim \Gamma(0,001;0,001)$ ;  $\sigma_\alpha \sim U(0,10000)$  e  $\sigma_\beta \sim U(0,10000)$ . Alguns autores afirmam que os resultados da análise Bayesiana em geral são insensíveis à priori não-informativa especificada (1;6), mas tem-se percebido que isso não é válido para os componentes de variância (6). A priori uniforme para o desvio tem sido considerada como uma boa opção, por isto esta foi a priori adotada (9).

O Amostrador de Gibbs foi implementado com 11000 iterações, sendo que as 1000 primeiras foram consideradas como período de aquecimento do algoritmo, em todas as replicações da simulação. A análise visual da trajetória da cadeia foi usada para verificar se esse período era suficiente para obter convergência.

Para cada conjunto, foram registradas as estimativas clássicas dos efeitos fixos, aleatórios e dos componentes de variância, e as estimativas a posteriori da média, moda e mediana para cada um dos parâmetros. A moda foi estimada através do método não paramétrico de suavização de *kernel* (“Kernel-Smoothing”). As estimativas obtidas foram comparadas com os valores verdadeiros dos parâmetros definidos inicialmente através do cálculo do vício relativo, definido por  $100 \frac{E(\hat{\theta}) - \theta}{\theta}$ , do erro quadrático médio, calculado como  $Var(\hat{\theta}) + [E(\hat{\theta}) - \theta]^2$ , e através da análise visual de box-plots. O programa R foi utilizado nas simulações e análises.

## Resultados e Discussões

Na análise visual da trajetória da cadeia, a exclusão das primeiras 1000 observações revelou que a cadeia seguiu um padrão similar ao longo das iterações, e houve convergência do algoritmo (resultados não mostrados). O Quadro 1 resume os resultados obtidos para as três configurações, mostrando os vieses relativos e erros quadráticos médios para todos os efeitos fixos e componentes de variância. A Figura 1 mostra a distribuição das estimativas para cada um dos componentes de variância do modelo para a configuração 1. A linha horizontal representa o verdadeiro valor do parâmetro, e o ponto representa a média. Analogamente, a Figura 2 e a Figura 3 apresentam os mesmos gráficos para as configurações 2 e 3, respectivamente.

Na maioria dos casos, ao menos onde as discrepâncias são relevantes, maiores valores de EQM estão associados com valores maiores do vício relativo. Isto era esperado devido à relação que essas medidas têm em sua definição. Nas três configurações, observou-se vício e EQM para os efeitos fixos e aleatórios de magnitude similar entre todos os estimadores estudados. Pelo fato de que as

estimativas dos efeitos aleatórios não são o principal interesse na prática, e são muitos parâmetros, optamos por omitir os resultados referentes a eles. Para o componente de variância  $\sigma_\varepsilon^2$ , novamente todos os estimadores apresentaram erro quadrático médio muito similar, o que pode ser visualizado através dos box-plots. Já em termos de viés, para este parâmetro, os quatro estimadores não apresentaram diferenças muito grandes. Para todas as configurações, o estimador mais viesado foi a moda, e o menos viesado, a média.

Nas três configurações simuladas, considerando a medida relativa do vício, alguns parâmetros mostraram-se marcadamente superestimados ou subestimados, e os vieses são na mesma direção para todos os métodos. Uma exceção ocorreu para o componente de variância  $\sigma_\alpha^2$ , cujos estimadores apresentam comportamentos diferentes. Na configuração 1, o estimador moda Bayesiana apresentou viés relativo (+18,70%) menor do que para os estimadores Bayesianos média (+268,34%), mediana (+174,97%) e para o estimador clássico (+143,47%). No entanto, o estimador moda Bayesiana subestima este componente de variância nas configurações 2 e 3, apresentando viés relativo de -39,03% e -34,95%, respectivamente.

As maiores diferenças encontradas entre os métodos foram em relação aos componentes de variância  $\sigma_\alpha^2$  e  $\sigma_\beta^2$ . Na configuração 1, para estes dois parâmetros, o menor vício e menor EQM foram apresentados pela moda Bayesiana. Isto pode ser visualizado na Figura 1, já que a moda foi o estimador que apresentou a menor dispersão. Nas Configurações 2 e 3, para o parâmetro  $\sigma_\alpha^2$ , o menor vício foi apresentado pelo estimador clássico, porém a moda Bayesiana mostrou menor EQM e segundo menor vício, sendo que a diferença entre os vieses não foi grande. Novamente, as Figuras 2 e 3 mostram que a distribuição das estimativas produzidas pelo estimador moda tem a menor variabilidade. Já para  $\sigma_\beta^2$ , o menor vício e menor EQM foram obtidos pela média Bayesiana e a moda apresentou maior vício e EQM. Esta contradição não permite concluir qual estimador é melhor, pois os estimadores média e moda se alternam entre o pior e o melhor estimador, dependendo da configuração e do parâmetro considerado.

É interessante notar que, em geral, todos os estimadores tiveram performance melhor na configuração 2 (onde há desbalanceamento em relação ao tempo) do que na configuração 1 (balanceada). Pode-se especular que, talvez, isto ocorra de forma similar ao modelo de regressão linear, onde quanto maior a variabilidade do preditor menores são as variâncias dos estimadores (10). Em sendo verdadeira essa relação, poderia trazer uma vantagem prática para o pesquisador, pois não haveria a necessidade de realizar as medidas em ocasiões igualmente espaçadas no tempo. Entretanto, é uma questão que deve ser pesquisada com maior profundidade.

A Figura 4 mostra as distribuições a posteriori dos componentes de variância estimadas não-parametricamente, para o conjunto de dados simulado na última replicação. É interessante notar que as distribuições para o componente de variância  $\sigma_\alpha^2$  são bastante assimétricas, e essa assimetria é maior com desbalanceamento e amostra pequena (configurações 1 e 2). É necessário, entanto, investigar com maior profundidade este comportamento.

Os resultados deste estudo são consistentes com aqueles obtidos anteriormente em comparações de modelos mistos para dados longitudinais (8), bem como com aspectos conceituais (1) e estudos de simulação para modelos mistos em outros contextos (5-7). Cabe salientar que, ao analisar o modelo de ANOVA com um efeito aleatório, Browne e Draper (5;6) consideraram priori uniforme para as variâncias, e não para o desvio-padrão, como considerado neste trabalho. Mesmo assim, a principal diferença observada entre os métodos pelos referidos autores também foi com relação ao componente de variância do efeito aleatório. Eles também identificaram boas propriedades da moda para estimar o componente de variância do efeito aleatório. Já o modelo de regressão com coeficientes aleatórios abordado por Browne e Draper (5;7) é diferente do modelo considerado neste estudo, pois não contempla covariáveis e assume efeitos aleatórios correlacionados. Além disso, as prioris utilizadas foram diferentes, e as configurações de tamanho de amostra simuladas pelos autores raramente seriam encontradas em um estudo de dados longitudinais. Mesmo com essas diferenças, os autores também encontraram maiores discrepâncias entre os métodos na estimação dos componentes de variância, reportando apenas resultados da média Bayesiana.

## Conclusões

Os resultados das simulações sugerem que a maior diferença entre as duas abordagens, no caso do modelo estudado, está na estimação dos componentes de variância. Nas configurações estudadas, pelo menos um dos estimadores Bayesianos apresentou menor EQM que o estimador clássico.

A moda Bayesiana mostrou-se um estimador com boas propriedades (menor vício relativo e menor EQM) para algumas situações, recomendando-se sua implementação no WinBUGS, que é o programa mais conhecido para análise através do Amostrador de Gibbs.

Um fato interessante foi que todos os estimadores tiveram melhor desempenho na situação desbalanceada em relação ao tempo, ou seja, quando os intervalos entre as medições não eram fixos. Entretanto, mais estudos são necessários para ter conclusões definitivas sobre o tópico.

Não é possível recomendar fortemente qual dos estimadores Bayesianos utilizar, pois o tópico exige mais estudos. O estudo de simulação poderia ser ampliado de várias maneiras. A primeira seria comparar, além da estimativa pontual que os métodos produzem para os vários parâmetros, a performance dos intervalos de confiança e credibilidade produzidos. Uma segunda maneira seria considerar diferentes valores para os parâmetros do modelo, e diferentes configurações de tamanho de amostra e desbalanceamento. Também seria interessante avaliar o comportamento dos estimadores considerando mais níveis na variável tratamento e maior número de covariáveis, bem como na presença de covariáveis contínuas. Outros modelos poderiam ser considerados como, por exemplo, uma relação quadrática entre tempo e a variável resposta, ou então uma distribuição multivariada para os efeitos aleatórios.

## Referências

- (1) Brown H, Prescott R. *Applied Mixed Models in Medicine*. Chichester: John Wiley & Sons, Inc.; 1999.
- (2) Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Second Edition ed. Boca Raton: Chapman & Hall; 2000.
- (3) Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall; 1996.
- (4) Spiegelhalter DJ, Thomas A, Best N, Gilks WR. *Bayesian Inference Using Gibbs Sampling: Version 0.5*. Cambridge, Inglaterra: Biostatistics Unit-MRC; 1995.
- (5) Browne WJ. *Applying MCMC Methods to Multilevel Models* Department of Mathematical Sciences, University of Bath, UK.; 1998.
- (6) Browne WJ, Draper D. A comparison of Bayesian and likelihood methods for fitting multilevel models. *Bayesian Analysis* 2006;1(3):473-514.
- (7) Browne WJ, Draper D. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics* 2000;15(3):391-420.
- (8) Leotti VB, Ziegelmann PK. *Abordagem Bayesiana para Modelos Mistos com o Uso de "Gibbs Sampling"* [Monografia para obtenção do título de Bacharel em Estatística.] Departamento de Estatística - UFRGS; 2005.
- (9) Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006;1(3):515-34.
- (10) Kutner MH. *Applied linear statistical models*. 5th ed. Boston: McGraw-Hill Irwin; 2005.

**Legendas:**

Figura 1 - Box-plots para os componentes de variância  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  e  $\sigma_\varepsilon^2$  na configuração 1.

Figura 2 - Box-plots para os componentes de variância  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  e  $\sigma_\varepsilon^2$  na configuração 2.

Figura 3 - Box-plots para os componentes de variância  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  e  $\sigma_\varepsilon^2$  na configuração 3.

Figura 4 - Distribuições a posteriori dos componentes de variância estimadas não-parametricamente para um dos conjuntos de dados gerados

Quadro 1 – Vício relativo e erro-quadrático médio para as três configurações

Figura 1

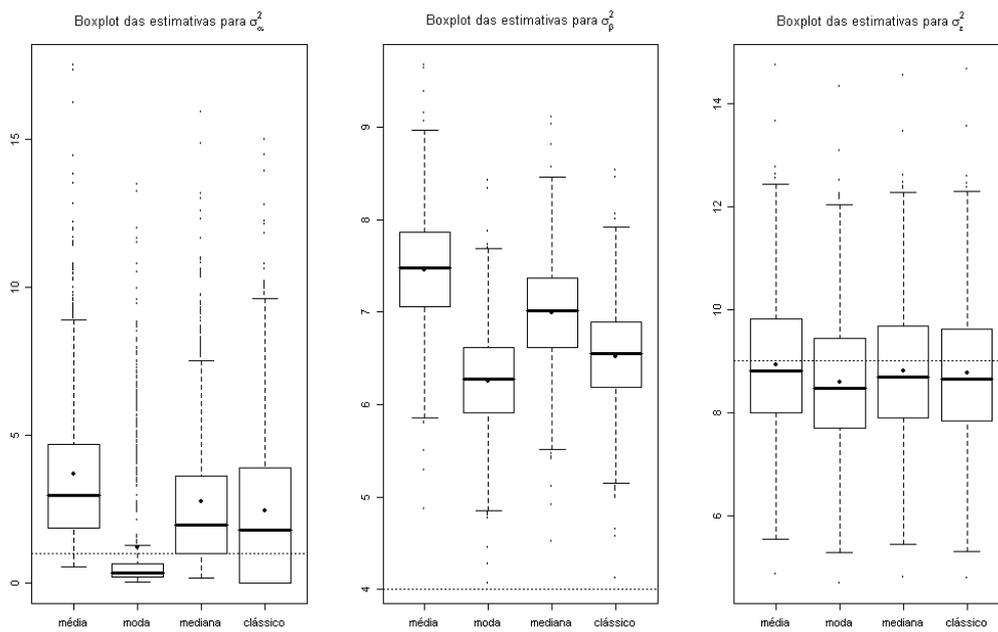


Figura 2

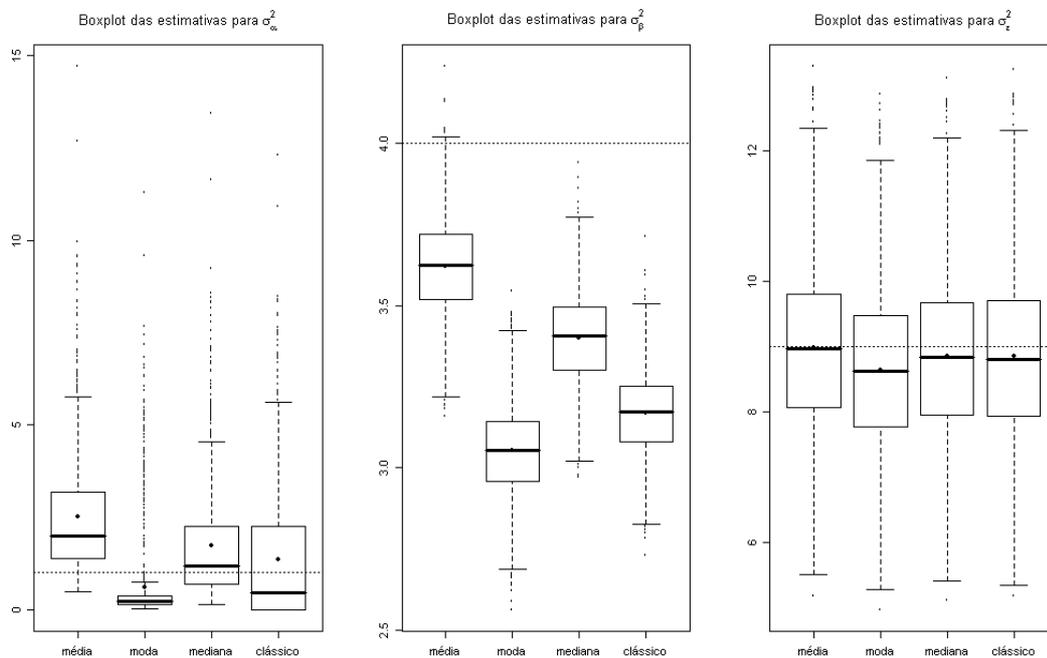


Figura 3

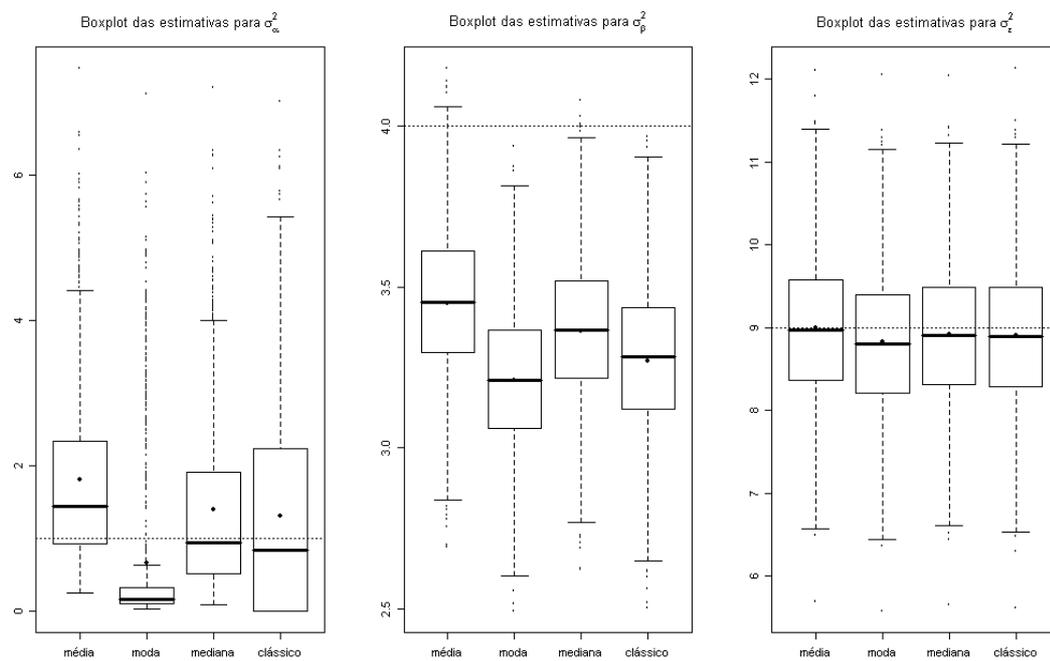
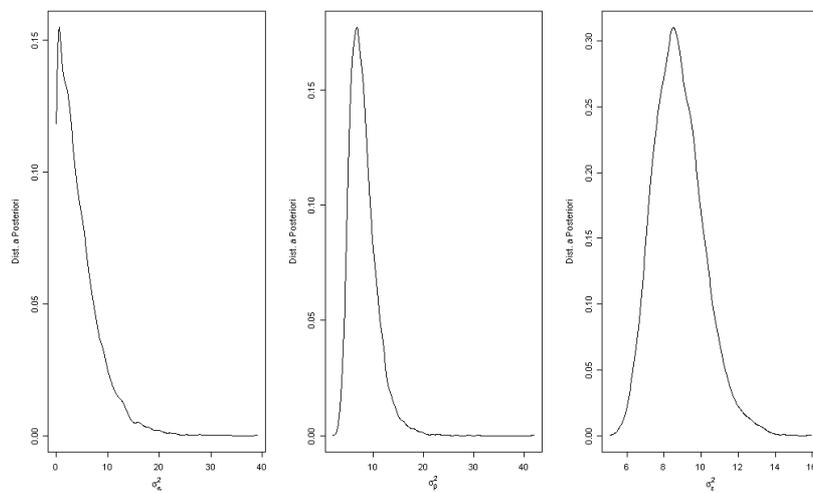
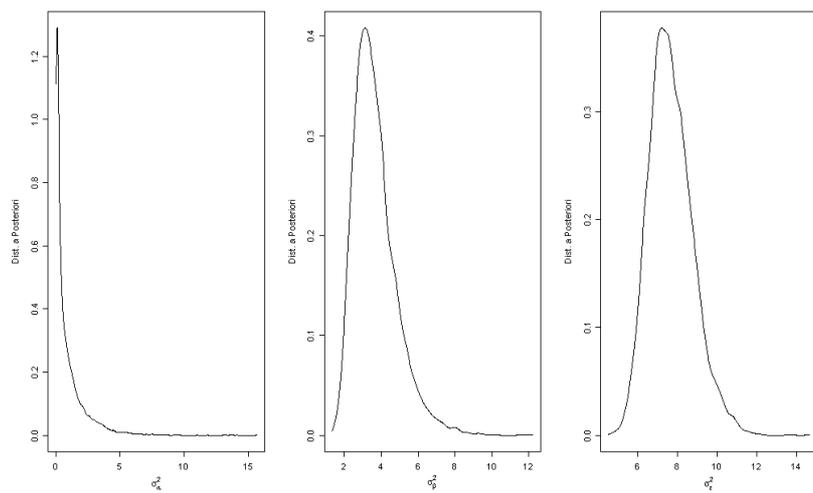


Figura 4

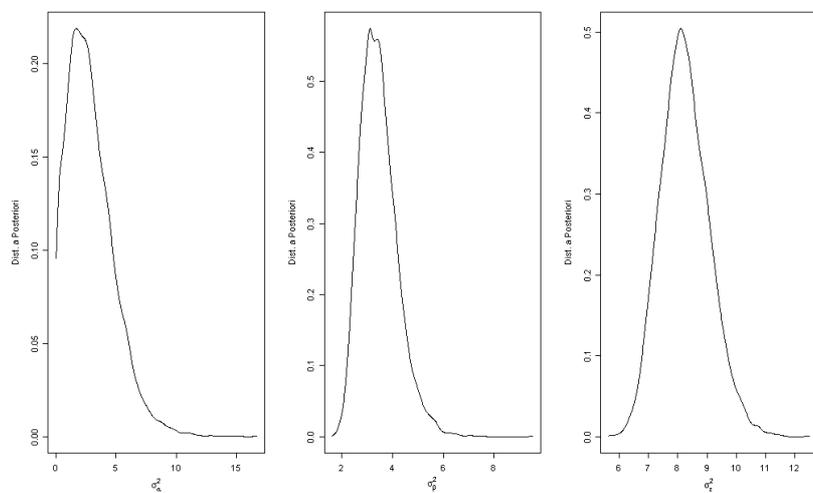
Configuração 1



Configuração 2



Configuração 3



Quadro 1

Config.	Parâmetro	Vício Relativo (%)				EQM			
		Média	Moda	Mediana	Clássico	Média	Moda	Mediana	Clássico
1	$\alpha_{trat1}$	4,60	4,59	4,59	4,61	1,33	1,44	1,34	1,33
	$\beta_{trat1}$	-15,80	-15,99	-15,82	-15,78	0,21	0,22	0,21	0,21
	$\alpha_{trat2}$	-4,86	-4,54	-4,86	-4,79	1,00	1,08	1,01	0,99
	$\beta_{trat2}$	5,08	4,97	5,07	5,11	0,13	0,14	0,13	0,13
	$\alpha_{trat3}$	23,61	24,19	23,92	23,62	1,17	1,25	1,17	1,14
	$\beta_{trat3}$	41,99	42,12	42,00	42,08	0,81	0,82	0,81	0,81
	$\sigma_{\alpha}^2$	268,34	18,70	174,97	143,47	13,60	4,71	9,00	9,29
	$\sigma_{\beta}^2$	86,37	56,32	74,74	63,10	12,32	5,38	9,28	6,69
	$\sigma_{\varepsilon}^2$	-0,68	-4,55	-2,02	-2,61	1,74	1,80	1,73	1,80
2	$\alpha_{trat1}$	7,97	8,11	7,97	8,05	1,24	1,33	1,24	1,24
	$\beta_{trat1}$	0,02	0,06	0,02	0,03	0,01	0,02	0,01	0,01
	$\alpha_{trat2}$	1,31	1,23	1,35	1,46	0,76	0,81	0,76	0,76
	$\beta_{trat2}$	-12,18	-12,21	-12,18	-12,16	0,25	0,25	0,25	0,25
	$\alpha_{trat3}$	11,10	11,06	11,32	11,65	1,01	1,13	1,02	1,00
	$\beta_{trat3}$	-14,16	-14,21	-14,17	-14,14	0,10	0,10	0,10	0,10
	$\sigma_{\alpha}^2$	151,38	-39,03	72,94	33,90	5,11	1,70	3,02	3,36
	$\sigma_{\beta}^2$	-9,46	-23,66	-14,95	-20,77	0,17	0,92	0,38	0,71
	$\sigma_{\varepsilon}^2$	-0,28	-4,04	-1,61	-1,72	1,76	1,78	1,74	1,80
3	$\alpha_{trat1}$	0,93	1,24	1,15	1,25	0,52	0,56	0,52	0,51
	$\beta_{trat1}$	13,29	13,59	13,44	13,57	0,12	0,12	0,12	0,12
	$\alpha_{trat2}$	-7,03	-6,73	-6,75	-6,71	0,59	0,61	0,57	0,57
	$\beta_{trat2}$	-15,41	-15,35	-15,34	-15,30	0,43	0,42	0,42	0,42
	$\alpha_{trat3}$	7,19	5,08	6,49	6,44	0,48	0,55	0,48	0,48
	$\beta_{trat3}$	-2,60	-2,57	-2,52	-2,48	0,05	0,05	0,05	0,05
	$\sigma_{\alpha}^2$	80,82	-34,96	39,06	30,35	2,13	1,53	1,64	2,20
	$\sigma_{\beta}^2$	-13,78	-19,76	-15,94	-18,23	0,36	0,68	0,46	0,59
	$\sigma_{\varepsilon}^2$	-0,05	-1,95	-0,89	-1,00	0,80	0,78	0,77	0,81

## 6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

Nesta dissertação os métodos MCMC foram introduzidos, destacando sua eficiência para resolver problemas de modelagem Bayesianas. Enfoque maior foi dado ao algoritmo Amostrador de Gibbs.

A abordagem clássica para os modelos mistos também foi revisada, devendo-se salientar que este modelo é muito útil para problemas estatísticos de modelagens mais sofisticadas, como o caso dos dados longitudinais. Além disso, ele unifica a análise de vários modelos segundo uma única abordagem.

Conceitos Bayesianos foram apresentados e aplicados para o contexto de modelo misto. A maior dificuldade da abordagem Bayesiana para os modelos mistos é que a distribuição a posteriori não pode ser derivada analiticamente, mas pode ser superada com o uso do Amostrador de Gibbs. Percebeu-se que a abordagem Bayesiana para modelos mistos apresenta vantagens teóricas sobre a abordagem clássica, pois as únicas suposições que são feitas, além das usuais do modelo, são em respeito às distribuições a priori. Já na abordagem clássica, além das suposições usuais de normalidade do erro e dos efeitos aleatórios, supõe-se que a matriz de variâncias seja conhecida para estimar os efeitos fixos e aleatórios. A suposição de uma distribuição a priori é bastante realista, pois é feita com base em toda a informação existente sobre o parâmetro e totalmente flexível à escolha do pesquisador. Por outro lado, a suposição de que as variâncias são conhecidas, necessária para estimar os efeitos fixos e aleatórios, raramente é atendida.

O estudo de simulação de dados longitudinais realizado sugere que a principal diferença entre as abordagens clássica e Bayesiana, no caso do modelo estudado, está na estimação dos componentes de variância. Em todas as configurações estudadas, pelo menos um dos estimadores Bayesianos apresentou menor EQM que o estimador clássico. A moda Bayesiana mostrou-se um estimador com boas propriedades para algumas situações, recomendando-se sua implementação no WinBUGS, que é o programa mais conhecido de Amostrador de Gibbs. Um fato interessante foi que todos os estimadores tiveram melhor desempenho na situação

desbalanceada em relação ao tempo. Entretanto, mais estudos são necessários para ter conclusões definitivas sobre o tópico.

Não é possível recomendar fortemente qual dos estimadores Bayesianos utilizar, pois o tópico exige mais estudos. O estudo de simulação poderia ser ampliado de várias maneiras. A primeira seria comparar, além da estimativa pontual que os métodos produzem para os vários parâmetros, a performance dos intervalos de confiança e credibilidade produzidos. Uma segunda maneira seria considerar diferentes valores para os parâmetros do modelo, e diferentes configurações de tamanho de amostra e desbalanceamento. Também seria interessante avaliar o comportamento dos estimadores considerando mais níveis na variável tratamento e maior número de covariáveis, bem como na presença de covariáveis contínuas. Outros modelos poderiam ser considerados como, por exemplo, uma relação quadrática entre tempo e a variável resposta, ou então uma distribuição multivariada para os efeitos aleatórios.

Como a abordagem Bayesiana se mostrou superior principalmente para a estimação dos componentes de variância, seria interessante estudar essa abordagem para os estudos de confiabilidade de medidas. Nesses estudos, alguns indivíduos são observados por alguns avaliadores um certo número de vezes, e se deseja avaliar a precisão das medições. O modelo utilizado para a análise desses dados é um modelo que pode ter somente efeitos aleatórios, ou ser misto. Uma das medidas mais utilizadas para avaliar o quanto o método de medição é reprodutível ou confiável é o coeficiente de correlação intraclassa (CCI), definido pela razão de componentes de variâncias. Neste sentido, seria interessante estudar e comparar os métodos Bayesiano e clássico para estimar o CCI.

Finalmente, o presente estudo mostrou que a técnica Bayesiana para dados longitudinais é mais precisa que a clássica principalmente na estimação dos componentes de variância. A sua aplicação prática ainda é um pouco prejudicada pelo fato de que o WinBUGS exige certo conhecimento de programação do usuário, ou seja, não é de tão fácil utilização como SAS e SPSS. Entretanto, muito material tem sido publicado, tanto na literatura como na Internet, o que facilita o aprendizado do assunto pelos pesquisadores.

## **7 ANEXOS**

### **7.1 ANEXO 1 – Projeto de Pesquisa**

### **7.2 ANEXO 2 – Códigos R para a Simulação**

### **7.3 ANEXO 3 – Exemplo de Gráfico de Convergência e posteriori estimada por Kernel**

### **7.4 ANEXO 4 – Descrição do WINBUGS**

## 7.1 ANEXO 1 – PROJETO DE PESQUISA

Universidade Federal do Rio Grande do Sul  
Faculdade de Medicina  
Programa de Pós-Graduação em Epidemiologia

Projeto de Pesquisa de Dissertação de Mestrado

# **Comparação por Simulação das Inferências Clássica e Bayesiana na Análise de Dados Longitudinais via Modelos Mistos**

*Vanessa Bielefeldt Leotti*

*Prof. Dr. Álvaro Vigo*

Porto Alegre, 29 de maio de 2007.

## ➤ Introdução

Dados longitudinais são medidas que são tomadas em mesmas unidades experimentais (indivíduos, animais, objetos, etc.) em diversos momentos ao longo do tempo. Este tipo de problema é bastante encontrado na área da Medicina, tanto em estudos observacionais como experimentais. Também se encontra dados longitudinais em experimentos agrônômicos e em estudos sobre desempenho educacional, entre outras áreas. Atualmente, a técnica mais utilizada para a análise de dados longitudinais é o ajuste de modelos mistos.

O modelo misto é um modelo que compreende efeitos fixos e aleatórios em sua formulação. Efeitos fixos são variáveis independentes que se supõe que assumam valores constantes que podem ser escolhidos arbitrariamente pelo pesquisador. Efeitos aleatórios são variáveis independentes que assumem valores considerados aleatórios, e supõe-se uma distribuição para estas variáveis. Outra característica do modelo misto é o fato de que se assume uma distribuição multivariada para os erros do modelo, o que permite analisar dados que não sejam independentes. Por isto, o modelo misto é bastante útil quando se tem dados que são agrupados de alguma forma, como, por exemplo, alunos agrupados em turmas, pacientes agrupados em hospitais, etc. Também devido a esta característica o modelo misto tem sido utilizado para analisar problemas envolvendo dados longitudinais, já que, para este tipo de dados, é coerente admitir a existência de correlação entre as observações feitas na mesma unidade. Nestes casos, a correlação inerente entre observações de um mesmo grupo pode ser incorporada no modelo. Normalmente, supõe-se distribuição normal multivariada para os erros e para os efeitos aleatórios, obtendo-se o modelo misto normal, que será o modelo focado neste trabalho.

Em notação matricial, podemos expressar o modelo misto como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

onde

$\mathbf{y}$  é um vetor  $n \times 1$  de dados observados;

$\mathbf{X}$  é uma matriz  $n \times p$  conhecida;

$\boldsymbol{\beta}$  é um vetor  $p \times 1$  de efeitos fixos geralmente desconhecidos;

$\mathbf{Z}$  é uma matriz  $n \times q$  conhecida;

$\mathbf{u}$  é um vetor  $q \times 1$  de efeitos aleatórios geralmente desconhecidos e  $\boldsymbol{\varepsilon}$  é um vetor  $n \times 1$  de erros aleatórios não-observáveis.

As suposições do modelo são:  $\mathbf{u} \sim N(0, \mathbf{D})$ ,  $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R})$  e  $Cov(\mathbf{u}, \boldsymbol{\varepsilon}') = 0$ , onde  $\mathbf{D}$  e  $\mathbf{R}$  são matrizes de variâncias e covariâncias de dimensões  $q \times q$  e  $n \times n$ , respectivamente.

Na análise de modelos mistos, pode-se supor um certo padrão para as matrizes de variâncias e covariâncias  $\mathbf{D}$  e  $\mathbf{R}$ . Isto é feito modelando-se uma estrutura de covariância para elas. Existem vários tipos de estruturas de covariância que podem ser modeladas, tais como as estruturas Componentes de Variância (permite modelar variâncias diferentes para grupos de efeitos ou observações e supõe covariâncias nulas) e Não-estruturada (permite variâncias e covariâncias totalmente diferentes).

Podemos analisar problemas envolvendo modelos mistos, assim como todos os outros problemas estatísticos, sob duas abordagens: clássica ou Bayesiana. A principal diferença entre a inferência clássica e a inferência Bayesiana é que a primeira supõe que os parâmetros desconhecidos do modelo analisado são constantes, enquanto a segunda os considera variáveis aleatórias. Sendo assim, a abordagem Bayesiana baseia-se em distribuições de probabilidade para os parâmetros desconhecidos. Antes de observar uma amostra de dados estas distribuições são construídas baseadas em toda a informação existente sobre estes parâmetros e são chamadas de distribuição a priori. A idéia da inferência Bayesiana é combinar a informação a priori com a informação proveniente dos dados amostrais. Ou seja, combinar a distribuição a priori e a função de verossimilhança, dando origem à distribuição a posteriori. Esta distribuição mostra o comportamento dos parâmetros após os dados amostrais serem observados e a partir dela são feitas inferências sobre os parâmetros e suas devidas interpretações. Quando não há informação disponível sobre algum dos parâmetros, pode-se utilizar uma priori não-informativa para este parâmetro, e então os resultados da análise Bayesiana serão muito similares aos resultados da análise clássica.

Para se fazer uma análise estatística através das técnicas Bayesianas para o modelo misto normal especificado acima, primeiramente precisa-se definir a distribuição a priori conjunta para os parâmetros desconhecidos, que são  $\boldsymbol{\beta}$ ,  $\mathbf{D}$  e  $\mathbf{R}$ .

Ou seja, precisa-se determinar a priori conjunta  $\pi(\boldsymbol{\beta}, \mathbf{D}, \mathbf{R})$ . Usualmente, supõe-se independência a priori entre estes parâmetros. Conseqüentemente,  $\pi(\boldsymbol{\beta}, \mathbf{D}, \mathbf{R}) = \pi(\boldsymbol{\beta})\pi(\mathbf{D})\pi(\mathbf{R})$ . Assumindo a independência a priori, necessita-se especificar distribuições para os parâmetros  $\boldsymbol{\beta}$ ,  $\mathbf{D}$  e  $\mathbf{R}$  separadamente. Deve-se notar que o vetor de efeitos aleatórios  $\mathbf{u}$  já tem sua distribuição de probabilidade especificada pelas suposições do modelo e por isso não se determina distribuição a priori para ele.

A distribuição a priori usual para  $\boldsymbol{\beta}$  é a distribuição Normal. As distribuições a priori para  $\mathbf{D}$  e  $\mathbf{R}$  dependerão da estrutura de covariância adotada para cada uma delas. Se a estrutura adotada for Componentes de Variância, usualmente adota-se distribuição a priori Gama para cada um dos parâmetros de variância da matriz. Se for admitida uma matriz de covariâncias Não-Estruturada, usualmente adota-se distribuição a priori Wishart.

Todo conjunto de dados pode ser analisado utilizando-se as técnicas Bayesianas. Entretanto, alguns problemas, como modelos com muitos parâmetros, a derivação da distribuição a posteriori exige cálculos muito difíceis, ou simplesmente impossíveis de serem executados. Este fato foi um empecilho para a utilização da inferência Bayesiana por muito tempo. Somente com o avanço dos métodos computacionais, especialmente dos métodos MCMC – “Markov Chain Monte Carlo”, que os métodos Bayesianos puderam ser aplicados a muitos problemas estatísticos, como o caso dos modelos mistos, que não possui distribuição a posteriori analiticamente derivável. O algoritmo de “Gibbs Sampling”, que é um dos métodos MCMC, fornece um método computacional eficaz para se obter estimativas da distribuição a posteriori para vários modelos, inclusive o modelo misto.

Na literatura, há poucos estudos que comparam as abordagens clássica e Bayesiana para modelos mistos sob um enfoque analítico. A maioria das discussões existentes aponta vantagens e desvantagens de cada um dos métodos apenas sob o ponto de vista filosófico. Alguns dos argumentos a favor da abordagem Bayesiana são:

- permite que a informação proveniente de outros estudos possa ser utilizada na análise do estudo atual, o que é impossível na inferência clássica;
- na maioria dos problemas, o número de suposições que devem ser feitas é

menor que na análise clássica, assim menos suposições devem ser testadas e a probabilidade de que o modelo não seja adequado diminui. Na inferência clássica, por exemplo, deve-se assumir distribuição para os estimadores, o que não ocorre na análise Bayesiana;

- no caso dos modelos mistos, os parâmetros são estimados por algoritmos iterativos que podem ter problemas com amostras pequenas, muitos dados faltantes, etc. A inferência Bayesiana geralmente se comporta melhor frente a essas situações;

- no caso dos parâmetros de variâncias e covariâncias do modelo misto, às vezes a análise clássica produz estimativas negativas de variâncias (método da ANOVA) ou então o método iterativo trunca e não fornece nenhuma estimativa (métodos ML e REML). Na inferência Bayesiana, geralmente esses problemas não ocorrem.

Browne (1) e Browne e Draper (2;3) procederam a estudos de simulação para alguns casos especiais do modelo misto e também do modelo misto generalizado, mas não abordaram modelos para dados longitudinais.

No contexto de dados longitudinais, um estudo de simulação (4), foi realizado para comparar os estimadores Bayesianos média, moda estimada não-parametricamente e mediana. Este estudo não contemplava a comparação com os estimadores clássicos. O modelo simulado considerou a estrutura de covariância Componentes de Variância para ambas as matrizes de covariância dos efeitos aleatórios e do erro. Além disso, considerou-se a situação de simular dados com uma estrutura de covariância Não-estruturada para a matriz de covariância do erro, e analisar estes dados supondo que a estrutura da mesma era Componentes de Variância. As conclusões do estudo foram as de que, ao supor erros independentes (ou seja, estrutura de covariância Componentes de Variância para o erro), quando na verdade os dados tinham erros correlacionados e com variâncias diferentes (ou seja, provenientes de uma estrutura de covariância Não-Estruturada para o erro), os estimadores dos efeitos fixos, e dos componentes de variância se tornam viesados. Além disto, conclui-se que a moda estimada não-parametricamente é um estimador com boas propriedades no caso do modelo estudado.

Por isso, o objetivo deste trabalho é elaborar um estudo de simulação de dados longitudinais, onde se possam comparar os dados analisados com ambas as

abordagens Bayesiana e clássica e avaliar qual delas tem as melhores propriedades em termos de vício e precisão.

### ➤ **Objetivos**

Proceder a um estudo de simulação de dados longitudinais para avaliar vício e precisão das estimativas obtidas através das análises clássica e Bayesiana para modelos mistos.

### ➤ **Métodos**

Para o estudo de simulação será considerado um modelo que envolve múltiplas medidas de um mesmo indivíduo (efeito de tempo) e uma única covariável que define grupos de indivíduos (efeito de tratamento). Serão escolhidos números de tempos, tratamentos e de indivíduos dentro de cada tratamento que representem situações de amostra pequena e desbalanceamento.

Serão determinados valores para os parâmetros do modelo e um conjunto de dados será gerado aleatoriamente conforme a estrutura do modelo. Assim, as análises clássica e Bayesiana serão procedidas e as estimativas obtidas serão registradas. Esse processo será repetido 1000 vezes e as estimativas registradas serão comparadas com os valores verdadeiros dos parâmetros definidos inicialmente. Essas comparações vão ser feitas através do cálculo do vício (média das estimativas obtidas menos o valor verdadeiro do parâmetro), do erro quadrático médio (variância das estimativas mais o quadrado do vício) e através da análise visual de box-plots. O software livre R será utilizado.

### ➤ **Questões Éticas**

Como o trabalho utiliza apenas dados simulados, não há restrições éticas envolvidas.

➤ **Recursos Necessários**

Os recursos necessários são material bibliográfico, um microcomputador e o software R, que é livre.

➤ **Cronograma**

Tarefa	2007					
	J	J	A	S	O	N
Redação da dissertação e artigo	■	■				
Defesa preliminar			■			
Correções				■		
Sessão pública					■	
Submissão para publicação						■

➤ **Bibliografia**

(1) Browne WJ. Applying MCMC Methods to Multilevel Models Department of Mathematical Sciences, University of Bath, UK.; 1998.

(2) Browne WJ, Draper D. A comparison of Bayesian and likelihood methods for fitting multilevel models. *Bayesian Analysis* 2006;1(3):473-514.

(3) Browne WJ, Draper D. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics* 2000;15(3):391-420.

(4) Leotti VB, Ziegelmann PK. Abordagem Bayesiana para Modelos Mistos com o Uso de "Gibbs Sampling" [Monografia para obtenção do título de Bacharel em Estatística.] Departamento de Estatística - UFRGS; 2005.

## 7.2 ANEXO 2 – CÓDIGOS R PARA A SIMULAÇÃO

Este código foi utilizado para a Configuração 1. O código das demais configurações é análogo.

```
# Precisa carregar o pacote nnet e nlme!

library(nnet)
library(nlme)

n = 27 #Número de sujeitos
g = 5 #Número de tempos
t = 3 #Número de tratamentos
q = 2 #número de efeitos aleatórios
p = 6 # número de efeitos fixos
cvar = 3 #Número de componentes de variância
nparms=p+q*n+cvar #Número de parâmetros

n1 = 8
n2 = 10
n3 = 9

trat = matrix(c(rep(1, n1), rep(0, (n-n1))), rep(0, n1), rep(1, n2), rep(0, n3)), nrow=n,
ncol=(t-1))

tempo = seq(1, 5, 1)
sum2tempo = sum(tempo^2)

alfa0.def = 1
alfa1.def = 5
alfa2.def = 3
beta0.def = 2
beta1.def = -4
beta2.def = -6

alfa.t1.def = alfa0.def+alfa1.def
alfa.t2.def = alfa0.def+alfa2.def
alfa.t3.def = alfa0.def
beta.t1.def = beta0.def+beta1.def
beta.t2.def = beta0.def+beta2.def
beta.t3.def = beta0.def

sigma2alfa.def = 1
sigma2beta.def = 4
sigma2e.def = 9

alfa.suj.def = rep(0,n)
beta.suj.def = rep(0,n)
```

```

for(i in 1:n) {
  alfa.suj.def[i] = rnorm(1, (alfa0.def+alfa1.def*trat[i,1]+alfa2.def*trat[i,2]),
sqrt(sigma2alfa.def))
  beta.suj.def[i] = rnorm(1, (beta0.def+beta1.def*trat[i,1]+beta2.def*trat[i,2]),
sqrt(sigma2beta.def))
}

# Parâmetros das prioris
a0 = 0
c02 = 10000
a1 = 0
c12 = 10000
a2 = 0
c22 = 10000
b0 = 0
d02 = 10000
b1 = 0
d12 = 10000
b2 = 0
d22 = 10000
e1 = 0
e2 = 0

# Gibbs Sampling
m = 11000 # Tamanho da cadeia
burn = 1000 # Valores a serem desprezados

rep = 1000 # Replicações da simulação

cred <- 95 # Percentual de credibilidade
alfa <- (1+cred/100)/2

media = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
mediana = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
moda = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
var = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc1 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc2 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
effix.class = matrix(rep(0, rep*p), nrow=rep, ncol=p)
alfa.suj.class = matrix(rep(0, rep*n), nrow=rep, ncol=n)
beta.suj.class = matrix(rep(0, rep*n), nrow=rep, ncol=n)
cvar.class = matrix(rep(0, rep*cvar), nrow=rep, ncol=cvar)

for(r in 1:rep){
  y = matrix(rep(0, n*g), nrow=n, ncol=g)
  yclass = matrix(rep(0, n*g*4), nrow=n*g, ncol=4)
  for(i in 1:n) {
    for(j in 1:g){
      y[i,j] = rnorm(1, (alfa.suj.def[i]+beta.suj.def[i]*tempo[j]),
sqrt(sigma2e.def))
    }
    yclass[(i+(i-1)*(g-1)):(i+i*(g-1)),1] = i
    if (trat[i,1]==1 & trat[i,2]==0) yclass[(i+(i-1)*(g-1)):(i+i*(g-1)),2] = 1
  }
}

```

```

else if (trat[i,1]==0 & trat[i,2]==1) yclass[(i+(i-1)*(g-1)):(i+i*(g-
1)),2] = 2
else yclass[(i+(i-1)*(g-1)):(i+i*(g-1)),2] = 3
yclass[(i+(i-1)*(g-1)):(i+i*(g-1)),3] = tempo
yclass[(i+(i-1)*(g-1)):(i+i*(g-1)),4] = y[i, ]
}

colnames(yclass) = c("Suj", "Trat", "Tempo", "Resp")
Dados = data.frame(yclass)
Dados[,1] = as.factor(Dados[,1])
Dados[,2] = as.factor(Dados[,2])

lme1 <- lme(Resp ~ Tempo*Trat, random = list(Suj=pdDiag(~ Tempo )),
data=Dados)

effix.class[r, ] = fixed.effects(lme1)
re = random.effects(lme1)
alfa.suj.class[r, ] = re[,1]
beta.suj.class[r, ] = re[,2]
vc = VarCorr(lme1)
cvar.class[r, ] = as.numeric(vc[,1])

alfa.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
beta.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
alfa0 = rep(0.01,m)
alfa1 = rep(0.01,m)
alfa2 = rep(0.01,m)
beta0 = rep(0.01,m)
beta1 = rep(0.01,m)
beta2 = rep(0.01,m)
sigma2e = rep(0.001, m)
sigma2alfa = rep(0.001, m)
sigma2beta = rep(0.001, m)

for (l in 1:(m-1)) {
aux1 = (n-1)/2
aux2 = sum((alfa.suj[l, ]-(alfa0[l] + alfa1[l]*trat[,1] + alfa2[l]*trat[
,2]))^2)/2
sigma2alfa[l+1] = 1/rgamma(1, shape=aux1, rate=aux2)

aux3 = sum((beta.suj[l, ]-(beta0[l] + beta1[l]*trat[,1] + beta2[l]*trat[
,2]))^2)/2
sigma2beta[l+1] = 1/rgamma(1, shape=aux1, rate=aux3)

aux4 = n*g/2 + e1
aux5 = rep(0, n)
for(i in 1:n){
aux5[i] = sum((y[i, ] - (alfa.suj[l, i] + beta.suj[l,i]*tempo))^2)
}

aux6 = sum(aux5)/2 + e2

```

```

sigma2e[l+1] = 1/rgamma(1, shape=aux4, rate=aux6)

aux7 = (c02*(sum(alfa.suj[l, ])-alfa1[l]*n1-
alfa2[l]*n2)+a0*sigma2alfa[l+1])/(sigma2alfa[l+1] + n*c02)
aux8 = (sigma2alfa[l+1]*c02)/(sigma2alfa[l+1] + n*c02)
alfa0[l+1] = rnorm(1, aux7, sqrt(aux8))

aux9 = (c12*(sum(trat[,1]*alfa.suj[l, ])-
n1*alfa0[l+1])+sigma2alfa[l+1]*a1)/(n1*c12+sigma2alfa[l+1])
aux10 = (sigma2alfa[l+1]*c12)/(sigma2alfa[l+1] + n1*c12)
alfa1[l+1] = rnorm(1, aux9, sqrt(aux10))

aux11 = (c22*(sum(trat[,2]*alfa.suj[l, ])-
n2*alfa0[l+1])+sigma2alfa[l+1]*a2)/(n2*c22+sigma2alfa[l+1])
aux12 = (sigma2alfa[l+1]*c22)/(sigma2alfa[l+1] + n2*c22)
alfa2[l+1] = rnorm(1, aux11, sqrt(aux12))

aux13 = (d02*(sum(beta.suj[l, ])-beta1[l]*n1-
beta2[l]*n2)+b0*sigma2beta[l+1])/(sigma2beta[l+1] + n*d02)
aux14 = (sigma2beta[l+1]*d02)/(sigma2beta[l+1] + n*d02)
beta0[l+1] = rnorm(1, aux13, sqrt(aux14))

aux15 = (d12*(sum(trat[,1]*beta.suj[l, ])-
n1*beta0[l+1])+sigma2beta[l+1]*b1)/(n1*d12+sigma2beta[l+1])
aux16 = (sigma2beta[l+1]*d12)/(sigma2beta[l+1] + n1*d12)
beta1[l+1] = rnorm(1, aux15, sqrt(aux16))

aux17 = (d22*(sum(trat[,2]*beta.suj[l, ])-
n2*beta0[l+1])+sigma2beta[l+1]*b2)/(n2*d22+sigma2beta[l+1])
aux18 = (sigma2beta[l+1]*d22)/(sigma2beta[l+1] + n2*d22)
beta2[l+1] = rnorm(1, aux17, sqrt(aux18))

aux19 =
(sigma2alfa[l+1]*sigma2e[l+1])/(g*sigma2alfa[l+1]+sigma2e[l+1])
aux20 =
(sigma2beta[l+1]*sigma2e[l+1])/(sum2tempo*sigma2beta[l+1]+sigma2e[l+1])
aux21 = rep(0, n)
aux22 = rep(0, n)
for(i in 1:n){
  aux21[i] = (sigma2alfa[l+1]*sum(y[i, ]-
beta.suj[l,i]*tempo)+sigma2e[l+1]*(alfa0[l+1]+alfa1[l+1]*trat[i,1]+alfa2[l+1]*trat[i,2]))/(
g*sigma2alfa[l+1]+sigma2e[l+1])
  alfa.suj[l+1, i] = rnorm(1, aux21[i], sqrt(aux19))
  aux22[i] = (sigma2beta[l+1]*sum(tempo*(y[i, ]-alfa.suj[l+1,
i]))+sigma2e[l+1]*(beta0[l+1]+beta1[l+1]*trat[i,1]+beta2[l+1]*trat[i,2]))/(sum2tempo*si
gma2beta[l+1]+sigma2e[l+1])
  beta.suj[l+1, i] = rnorm(1, aux22[i], sqrt(aux20))
}
}

alfa.t1 = alfa0+alfa1
alfa.t2 = alfa0+alfa2
alfa.t3 = alfa0

```

```

beta.t1 = beta0+beta1
beta.t2 = beta0+beta2
beta.t3 = beta0

```

```

alfa.t1.class = effix.class[ ,1]
alfa.t2.class = effix.class[ ,1]+effix.class[ ,3]
alfa.t3.class = effix.class[ ,1]+effix.class[ ,4]
beta.t1.class = effix.class[ ,2]
beta.t2.class = effix.class[ ,2]+effix.class[ ,5]
beta.t3.class = effix.class[ ,2]+effix.class[ ,6]

```

```
# Resultados do Gibbs Sampling
```

```

for(i in 1:n) {
  media[r, i] = mean(alfa.suj[(burn+1):m, i])
  media[r, (i+n)] = mean(beta.suj[(burn+1):m, i])
  mediana[r, i] = median(alfa.suj[(burn+1):m, i])
  mediana[r, (i+n)] = median(beta.suj[(burn+1):m, i])
  perc1[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(1-alfa))
  perc2[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(alfa))
  perc1[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(1-alfa))
  perc2[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(alfa))
  var[r, i] = var(alfa.suj[(burn+1):m, i])
  var[r, (i+n)] = var(beta.suj[(burn+1):m, i])
  kernel1=density(alfa.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(alfa.suj[(burn+1):m, i]), to=max(alfa.suj[(burn+1):m, i]))
  posmo1 = which.is.max(kernel1$y)
  moda[r, i] = kernel1$x[posmo1]
  kernel2=density(beta.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(beta.suj[(burn+1):m, i]), to=max(beta.suj[(burn+1):m, i]))
  posmo2 = which.is.max(kernel2$y)
  moda[r, (i+n)] = kernel2$x[posmo2]
}

```

```

media[r, 55] = mean(alfa.t1[(burn+1):m])
media[r, 56] = mean(alfa.t2[(burn+1):m])
media[r, 57] = mean(alfa.t3[(burn+1):m])
media[r, 58] = mean(beta.t1[(burn+1):m])
media[r, 59] = mean(beta.t2[(burn+1):m])
media[r, 60] = mean(beta.t3[(burn+1):m])
media[r, 61] = mean(sigma2alfa[(burn+1):m])
media[r, 62] = mean(sigma2beta[(burn+1):m])
media[r, 63] = mean(sigma2e[(burn+1):m])

```

```

mediana[r, 55] = median(alfa.t1[(burn+1):m])
mediana[r, 56] = median(alfa.t2[(burn+1):m])
mediana[r, 57] = median(alfa.t3[(burn+1):m])
mediana[r, 58] = median(beta.t1[(burn+1):m])
mediana[r, 59] = median(beta.t2[(burn+1):m])
mediana[r, 60] = median(beta.t3[(burn+1):m])
mediana[r, 61] = median(sigma2alfa[(burn+1):m])
mediana[r, 62] = median(sigma2beta[(burn+1):m])
mediana[r, 63] = median(sigma2e[(burn+1):m])

```

```

perc1[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(1-alfa))
perc1[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(1-alfa))
perc1[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(1-alfa))
perc1[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(1-alfa))
perc1[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(1-alfa))
perc1[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(1-alfa))
perc1[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(1-alfa))
perc1[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(1-alfa))
perc1[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(1-alfa))

```

```

perc2[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(alfa))
perc2[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(alfa))
perc2[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(alfa))
perc2[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(alfa))
perc2[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(alfa))
perc2[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(alfa))
perc2[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(alfa))
perc2[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(alfa))
perc2[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(alfa))

```

```

var[r, 55] = var(alfa.t1[(burn+1):m])
var[r, 56] = var(alfa.t2[(burn+1):m])
var[r, 57] = var(alfa.t3[(burn+1):m])
var[r, 58] = var(beta.t1[(burn+1):m])
var[r, 59] = var(beta.t2[(burn+1):m])
var[r, 60] = var(beta.t3[(burn+1):m])
var[r, 61] = var(sigma2alfa[(burn+1):m])
var[r, 62] = var(sigma2beta[(burn+1):m])
var[r, 63] = var(sigma2e[(burn+1):m])

```

```

kernel3 = density(alfa.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t1[(burn+1):m]), to=max(alfa.t1[(burn+1):m]))
posmo3 = which.is.max(kernel3$y)
moda[r, 55] = kernel3$x[posmo3]

```

```

kernel4 = density(alfa.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t2[(burn+1):m]), to=max(alfa.t2[(burn+1):m]))
posmo4 = which.is.max(kernel4$y)
moda[r, 56] = kernel4$x[posmo4]

```

```

kernel5 = density(alfa.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t3[(burn+1):m]), to=max(alfa.t3[(burn+1):m]))
posmo5 = which.is.max(kernel5$y)
moda[r, 57] = kernel5$x[posmo5]

```

```

kernel6 = density(beta.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t1[(burn+1):m]), to=max(beta.t1[(burn+1):m]))
posmo6 = which.is.max(kernel6$y)
moda[r, 58] = kernel6$x[posmo6]

```

```

kernel7 = density(beta.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t2[(burn+1):m]), to=max(beta.t2[(burn+1):m]))

```

```

posmo7 = which.is.max(kernel7$y)
moda[r, 59] = kernel7$x[posmo7]

kernel8 = density(beta.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t3[(burn+1):m]), to=max(beta.t3[(burn+1):m]))
posmo8 = which.is.max(kernel8$y)
moda[r, 60] = kernel8$x[posmo8]

kernel9 = density(sigma2alfa[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2alfa[(burn+1):m]), to=max(sigma2alfa[(burn+1):m]))
posmo9 = which.is.max(kernel9$y)
moda[r, 61] = kernel9$x[posmo9]

kernel10 = density(sigma2beta[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2beta[(burn+1):m]), to=max(sigma2beta[(burn+1):m]))
posmo10 = which.is.max(kernel10$y)
moda[r, 62] = kernel10$x[posmo10]

kernel11 = density(sigma2e[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2e[(burn+1):m]), to=max(sigma2e[(burn+1):m]))
posmo11 = which.is.max(kernel11$y)
moda[r, 63] = kernel11$x[posmo11]

print(r)
}

# Corrigindo média dos efeitos aleatórios clássicos

alfa.suj.class.m = matrix(rep(0, rep*n), nrow=rep, ncol=n)
beta.suj.class.m = matrix(rep(0, rep*n), nrow=rep, ncol=n)

for (i in 1:n1) {
  alfa.suj.class.m[, i] = alfa.suj.class[, i]+alfa.t1.class
  beta.suj.class.m[, i] = beta.suj.class[, i]+beta.t1.class
}

for (i in 1:n2) {
  alfa.suj.class.m[, (n1+i)] = alfa.suj.class[, (n1+i)]+alfa.t2.class
  beta.suj.class.m[, (n1+i)] = beta.suj.class[, (n1+i)]+beta.t2.class
}

for (i in 1:n3) {
  alfa.suj.class.m[, (n1+n2+i)] = alfa.suj.class[, (n1+n2+i)]+alfa.t3.class
  beta.suj.class.m[, (n1+n2+i)] = beta.suj.class[, (n1+n2+i)]+beta.t3.class
}

# Gráficos de convergência

windows()
par(mfrow=c(1,3))
ts.plot(sigma2alfa)
ts.plot(sigma2beta)
ts.plot(sigma2e)

```

```

windows()
par(mfrow=c(2,3))
ts.plot(alfa.t1)
ts.plot(alfa.t2)
ts.plot(alfa.t3)
ts.plot(beta.t1)
ts.plot(beta.t2)
ts.plot(beta.t3)

```

```

windows()
par(mfrow=c(2,2))
ts.plot(alfa.suj[,1])
ts.plot(beta.suj[,1])
ts.plot(alfa.suj[,2])
ts.plot(beta.suj[,2])

```

# Box-plots e cálculos de vício e eqm

```

windows()
par(mfrow=c(3,2))

```

```

boxplot(media[ ,1], moda[ ,1], mediana[ ,1], alfa.suj.class.m[ ,1], names=c("média",
"moda", "mediana", "clássico"))
abline(h = alfa.suj.def[1], lty="dotted")
title(main="Boxplot das estimativas para alfa.suj.1")

```

```

v.alfa.suj.1.media = mean(media[ ,1]) - alfa.suj.def[1]
eqm.alfa.suj.1.media = var(media[ ,1]) + (v.alfa.suj.1.media)^2

```

```

v.alfa.suj.1.moda = mean(moda[ ,1]) - alfa.suj.def[1]
eqm.alfa.suj.1.moda = var(moda[ ,1]) + (v.alfa.suj.1.moda)^2

```

```

v.alfa.suj.1.mediana = mean(mediana[ ,1]) - alfa.suj.def[1]
eqm.alfa.suj.1.mediana = var(mediana[ ,1]) + (v.alfa.suj.1.mediana)^2

```

```

v.alfa.suj.1.class = mean(alfa.suj.class.m[ ,1]) - alfa.suj.def[1]
eqm.alfa.suj.1.class = var(alfa.suj.class.m[ ,1]) + (v.alfa.suj.1.class)^2

```

```

boxplot(media[ ,28], moda[ ,28], mediana[ ,28], beta.suj.class.m[ ,1],
names=c("média", "moda", "mediana", "clássico"))
abline(h = beta.suj.def[1], lty="dotted")
title(main="Boxplot das estimativas para beta.suj.1")

```

```

v.beta.suj.1.media = mean(media[ ,28]) - beta.suj.def[1]
eqm.beta.suj.1.media = var(media[ ,28]) + (v.beta.suj.1.media)^2

```

```

v.beta.suj.1.moda = mean(moda[ ,28]) - beta.suj.def[1]
eqm.beta.suj.1.moda = var(moda[ ,28]) + (v.beta.suj.1.moda)^2

```

```

v.beta.suj.1.mediana = mean(mediana[ ,28]) - beta.suj.def[1]
eqm.beta.suj.1.mediana = var(mediana[ ,28]) + (v.beta.suj.1.mediana)^2

```

```
v.beta.suj.1.class = mean(beta.suj.class.m[ ,1]) - beta.suj.def[1]
eqm.beta.suj.1.class = var(beta.suj.class.m[ ,1]) + (v.beta.suj.1.class)^2
```

```
boxplot(media[ ,9], moda[ ,9], mediana[ ,9], alfa.suj.class.m[ ,9], names=c("média",
"moda", "mediana", "clássico"))
abline(h = alfa.suj.def[9], lty="dotted")
title(main="Boxplot das estimativas para alfa.suj.9")
```

```
v.alfa.suj.9.media = mean(media[ ,9]) - alfa.suj.def[9]
eqm.alfa.suj.9.media = var(media[ ,9]) + (v.alfa.suj.9.media)^2
```

```
v.alfa.suj.9.moda = mean(moda[ ,9]) - alfa.suj.def[9]
eqm.alfa.suj.9.moda = var(moda[ ,9]) + (v.alfa.suj.9.moda)^2
```

```
v.alfa.suj.9.mediana = mean(mediana[ ,9]) - alfa.suj.def[9]
eqm.alfa.suj.9.mediana = var(mediana[ ,9]) + (v.alfa.suj.9.mediana)^2
```

```
v.alfa.suj.9.class = mean(alfa.suj.class.m[ ,9]) - alfa.suj.def[9]
eqm.alfa.suj.9.class = var(alfa.suj.class.m[ ,9]) + (v.alfa.suj.9.class)^2
```

```
boxplot(media[ ,36], moda[ ,36], mediana[ ,36], beta.suj.class.m[ ,9],
names=c("média", "moda", "mediana", "clássico"))
abline(h = beta.suj.def[9], lty="dotted")
title(main="Boxplot das estimativas para beta.suj.9")
```

```
v.beta.suj.9.media = mean(media[ ,36]) - beta.suj.def[9]
eqm.beta.suj.9.media = var(media[ ,36]) + (v.beta.suj.9.media)^2
```

```
v.beta.suj.9.moda = mean(moda[ ,36]) - beta.suj.def[9]
eqm.beta.suj.9.moda = var(moda[ ,36]) + (v.beta.suj.9.moda)^2
```

```
v.beta.suj.9.mediana = mean(mediana[ ,36]) - beta.suj.def[9]
eqm.beta.suj.9.mediana = var(mediana[ ,36]) + (v.beta.suj.9.mediana)^2
```

```
v.beta.suj.9.class = mean(beta.suj.class.m[ ,9]) - beta.suj.def[9]
eqm.beta.suj.9.class = var(beta.suj.class.m[ ,9]) + (v.beta.suj.9.class)^2
```

```
boxplot(media[ ,19], moda[ ,19], mediana[ ,19], alfa.suj.class.m[ ,19],
names=c("média", "moda", "mediana", "clássico"))
abline(h = alfa.suj.def[19], lty="dotted")
title(main="Boxplot das estimativas para alfa.suj.19")
```

```
v.alfa.suj.19.media = mean(media[ ,19]) - alfa.suj.def[19]
eqm.alfa.suj.19.media = var(media[ ,19]) + (v.alfa.suj.19.media)^2
```

```
v.alfa.suj.19.moda = mean(moda[ ,19]) - alfa.suj.def[19]
eqm.alfa.suj.19.moda = var(moda[ ,19]) + (v.alfa.suj.19.moda)^2
```

```
v.alfa.suj.19.mediana = mean(mediana[ ,19]) - alfa.suj.def[19]
```

```
eqm.alfa.suj.19.mediana = var(media[ ,19]) + (v.alfa.suj.19.mediana)^2
```

```
v.alfa.suj.19.class = mean(alfa.suj.class.m[ ,19]) - alfa.suj.def[19]
eqm.alfa.suj.19.class = var(alfa.suj.class.m[ ,19]) + (v.alfa.suj.19.class)^2
```

```
boxplot(media[ ,46], moda[ ,46], mediana[ ,46], beta.suj.class.m[ ,19],
names=c("média", "moda", "mediana", "clássico"))
abline(h = beta.suj.def[19], lty="dotted")
title(main="Boxplot das estimativas para beta.suj.19")
```

```
v.beta.suj.19.media = mean(media[ ,46]) - beta.suj.def[19]
eqm.beta.suj.19.media = var(media[ ,46]) + (v.beta.suj.19.media)^2
```

```
v.beta.suj.19.moda = mean(modas[ ,46]) - beta.suj.def[19]
eqm.beta.suj.19.moda = var(modas[ ,46]) + (v.beta.suj.19.moda)^2
```

```
v.beta.suj.19.mediana = mean(mediana[ ,46]) - beta.suj.def[19]
eqm.beta.suj.19.mediana = var(mediana[ ,46]) + (v.beta.suj.19.mediana)^2
```

```
v.beta.suj.19.class = mean(beta.suj.class.m[ ,19]) - beta.suj.def[19]
eqm.beta.suj.19.class = var(beta.suj.class.m[ ,19]) + (v.beta.suj.19.class)^2
```

```
windows()
par(mfrow=c(3,2))
```

```
boxplot(media[ ,55], moda[ ,55], mediana[ ,55], alfa.t1.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = alfa.t1.def, lty="dotted")
title(main="Boxplot das estimativas para alfa.t1")
```

```
v.alfa.t1.media = mean(media[ ,55]) - alfa.t1.def
eqm.alfa.t1.media = var(media[ ,55]) + (v.alfa.t1.media)^2
```

```
v.alfa.t1.moda = mean(modas[ ,55]) - alfa.t1.def
eqm.alfa.t1.moda = var(modas[ ,55]) + (v.alfa.t1.moda)^2
```

```
v.alfa.t1.mediana = mean(mediana[ ,55]) - alfa.t1.def
eqm.alfa.t1.mediana = var(mediana[ ,55]) + (v.alfa.t1.mediana)^2
```

```
v.alfa.t1.class = mean(alfa.t1.class) - alfa.t1.def
eqm.alfa.t1.class = var(alfa.t1.class) + (v.alfa.t1.class)^2
```

```
boxplot(media[ ,58], moda[ ,58], mediana[ ,58], beta.t1.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = beta.t1.def, lty="dotted")
title(main="Boxplot das estimativas para beta.t1")
```

```
v.beta.t1.media = mean(media[ ,58]) - beta.t1.def
eqm.beta.t1.media = var(media[ ,58]) + (v.beta.t1.media)^2
```

```
v.beta.t1.moda = mean(modas[,58]) - beta.t1.def
eqm.beta.t1.moda = var(modas[,58]) + (v.beta.t1.moda)^2
```

```
v.beta.t1.mediana = mean(mediana[,58]) - beta.t1.def
eqm.beta.t1.mediana = var(mediana[,58]) + (v.beta.t1.mediana)^2
```

```
v.beta.t1.class = mean(beta.t1.class) - beta.t1.def
eqm.beta.t1.class = var(beta.t1.class) + (v.beta.t1.class)^2
```

```
boxplot(media[,56], moda[,56], mediana[,56], alfa.t2.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = alfa.t2.def, lty="dotted")
title(main="Boxplot das estimativas para alfa.t2")
```

```
v.alfa.t2.media = mean(media[,56]) - alfa.t2.def
eqm.alfa.t2.media = var(media[,56]) + (v.alfa.t2.media)^2
```

```
v.alfa.t2.moda = mean(modas[,56]) - alfa.t2.def
eqm.alfa.t2.moda = var(modas[,56]) + (v.alfa.t2.moda)^2
```

```
v.alfa.t2.mediana = mean(mediana[,56]) - alfa.t2.def
eqm.alfa.t2.mediana = var(mediana[,56]) + (v.alfa.t2.mediana)^2
```

```
v.alfa.t2.class = mean(alfa.t2.class) - alfa.t2.def
eqm.alfa.t2.class = var(alfa.t2.class) + (v.alfa.t2.class)^2
```

```
boxplot(media[,59], moda[,59], mediana[,59], beta.t2.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = beta.t2.def, lty="dotted")
title(main="Boxplot das estimativas para beta.t2")
```

```
v.beta.t2.media = mean(media[,59]) - beta.t2.def
eqm.beta.t2.media = var(media[,59]) + (v.beta.t2.media)^2
```

```
v.beta.t2.moda = mean(modas[,59]) - beta.t2.def
eqm.beta.t2.moda = var(modas[,59]) + (v.beta.t2.moda)^2
```

```
v.beta.t2.mediana = mean(mediana[,59]) - beta.t2.def
eqm.beta.t2.mediana = var(mediana[,59]) + (v.beta.t2.mediana)^2
```

```
v.beta.t2.class = mean(beta.t2.class) - beta.t2.def
eqm.beta.t2.class = var(beta.t2.class) + (v.beta.t2.class)^2
```

```
boxplot(media[,57], moda[,57], mediana[,57], alfa.t3.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = alfa.t3.def, lty="dotted")
title(main="Boxplot das estimativas para alfa.t3")
```

```
v.alfa.t3.media = mean(media[,57]) - alfa.t3.def
eqm.alfa.t3.media = var(media[,57]) + (v.alfa.t3.media)^2
```

```
v.alfa.t3.moda = mean(modas[,57]) - alfa.t3.def
eqm.alfa.t3.moda = var(modas[,57]) + (v.alfa.t3.moda)^2
```

```
v.alfa.t3.mediana = mean(mediana[,57]) - alfa.t3.def
eqm.alfa.t3.mediana = var(mediana[,57]) + (v.alfa.t3.mediana)^2
```

```
v.alfa.t3.class = mean(alfa.t3.class) - alfa.t3.def
eqm.alfa.t3.class = var(alfa.t3.class) + (v.alfa.t3.class)^2
```

```
boxplot(media[,60], modas[,60], mediana[,60], beta.t3.class, names=c("média",
"moda", "mediana", "clássico"))
abline(h = beta.t3.def, lty="dotted")
title(main="Boxplot das estimativas para beta.t3")
```

```
v.beta.t3.media = mean(media[,60]) - beta.t3.def
eqm.beta.t3.media = var(media[,60]) + (v.beta.t3.media)^2
```

```
v.beta.t3.moda = mean(modas[,60]) - beta.t3.def
eqm.beta.t3.moda = var(modas[,60]) + (v.beta.t3.moda)^2
```

```
v.beta.t3.mediana = mean(mediana[,60]) - beta.t3.def
eqm.beta.t3.mediana = var(mediana[,60]) + (v.beta.t3.mediana)^2
```

```
v.beta.t3.class = mean(beta.t3.class) - beta.t3.def
eqm.beta.t3.class = var(beta.t3.class) + (v.beta.t3.class)^2
```

```
windows()
par(mfrow=c(3,1))
```

```
boxplot(media[,61], modas[,61], mediana[,61], cvar.class[,1], names=c("média",
"moda", "mediana", "clássico"))
abline(h = sigma2alfa.def, lty="dotted")
title(main="Boxplot das estimativas para sigma2alfa")
```

```
v.sigma2alfa.media = mean(media[,61]) - sigma2alfa.def
eqm.sigma2alfa.media = var(media[,61]) + (v.sigma2alfa.media)^2
```

```
v.sigma2alfa.moda = mean(modas[,61]) - sigma2alfa.def
eqm.sigma2alfa.moda = var(modas[,61]) + (v.sigma2alfa.moda)^2
```

```
v.sigma2alfa.mediana = mean(mediana[,61]) - sigma2alfa.def
eqm.sigma2alfa.mediana = var(mediana[,61]) + (v.sigma2alfa.mediana)^2
```

```
v.sigma2alfa.class = mean(cvar.class[,1]) - sigma2alfa.def
eqm.sigma2alfa.class = var(cvar.class[,1]) + (v.sigma2alfa.class)^2
```

```
boxplot(media[,62], modas[,62], mediana[,62], cvar.class[,2], names=c("média",
"moda", "mediana", "clássico"))
abline(h = sigma2beta.def, lty="dotted")
```

```

title(main="Boxplot das estimativas para sigma2beta")

v.sigma2beta.media = mean(media[ ,62]) - sigma2beta.def
eqm.sigma2beta.media = var(media[ ,62]) + (v.sigma2beta.media)^2

v.sigma2beta.moda = mean(modas[ ,62]) - sigma2beta.def
eqm.sigma2beta.moda = var(modas[ ,62]) + (v.sigma2beta.moda)^2

v.sigma2beta.mediana = mean(mediana[ ,62]) - sigma2beta.def
eqm.sigma2beta.mediana = var(mediana[ ,62]) + (v.sigma2beta.mediana)^2

v.sigma2beta.class = mean(cvar.class[ ,2]) - sigma2beta.def
eqm.sigma2beta.class = var(cvar.class[ ,2]) + (v.sigma2beta.class)^2

boxplot(media[ ,63], modas[ ,63], mediana[ ,63], cvar.class[ ,3], names=c("média",
"moda", "mediana", "clássico"))
abline(h = sigma2e.def, lty="dotted")
title(main="Boxplot das estimativas para sigma2e")

v.sigma2e.media = mean(media[ ,63]) - sigma2e.def
eqm.sigma2e.media = var(media[ ,63]) + (v.sigma2e.media)^2

v.sigma2e.moda = mean(modas[ ,63]) - sigma2e.def
eqm.sigma2e.moda = var(modas[ ,63]) + (v.sigma2e.moda)^2

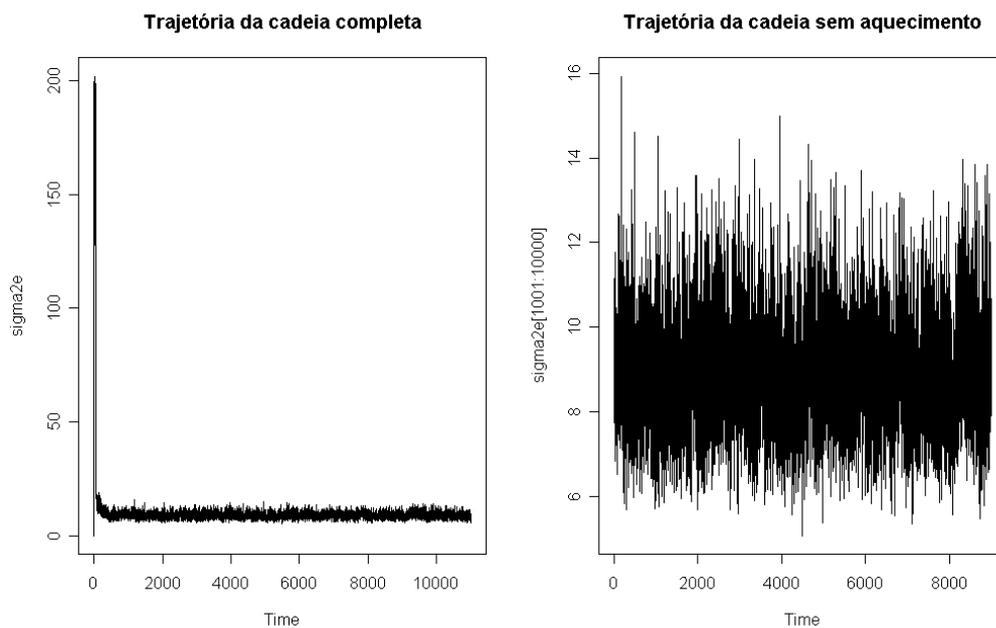
v.sigma2e.mediana = mean(mediana[ ,63]) - sigma2e.def
eqm.sigma2e.mediana = var(mediana[ ,63]) + (v.sigma2e.mediana)^2

v.sigma2e.class = mean(cvar.class[ ,3]) - sigma2e.def
eqm.sigma2e.class = var(cvar.class[ ,3]) + (v.sigma2e.class)^2

```

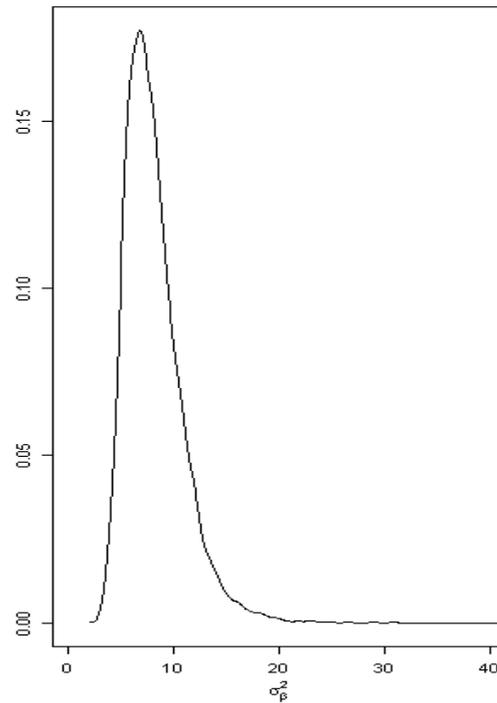
### 7.3 ANEXO 3 – EXEMPLO DE GRÁFICO DE CONVERGÊNCIA E DE POSTERIORI ESTIMADA POR KERNEL

No estudo de simulação, fez-se gráficos como os abaixo para verificar a convergência do algoritmo “Gibbs Sampling”:



Ao analisar o gráfico da cadeia completa, percebe-se que as primeiras iterações foram muito discrepantes das demais. Em geral, a partir da 1001ª iteração, a cadeia parecia ter estabelecido um padrão de comportamento. Para verificar isso, se desconsiderou as primeiras 1000 observações e se fez novamente o gráfico da trajetória. Parece que a cadeia seguiu um padrão por todas as demais iterações, por isso acredita-se que o período de 1000 iterações foi suficiente para a convergência da cadeia para esse parâmetro. Para os demais parâmetros, a análise foi similar.

Também no estudo de simulação, para um particular conjunto de dados gerado, e para o parâmetro  $\sigma_{\beta}^2$ , estimou-se por suavização de Kernel a seguinte distribuição a posteriori.



Observa-se que a posteriori para este parâmetro é assimétrica, fato que é comum quando se trata de variâncias. Observa-se que há muito pouca probabilidade desse parâmetro ser um valor maior que 20. O valor máximo a posteriori (moda) é próximo a 10, que foi o verdadeiro valor definido na simulação.

## 7.4 ANEXO 4 – DESCRIÇÃO DO WINBUGS

O programa WinBUGS está disponível gratuitamente na página <http://www.mrc-bsu.cam.ac.uk/bugs/> e atualmente está na versão 1.4.3. Ao instalá-lo, é necessário fazer um registro e obter uma chave de acesso para poder se utilizar todos os recursos do software. Este registro é fácil de ser feito, seguindo as instruções da página, e tem apenas a finalidade de prover informações aos idealizadores e patrocinadores do software de quantas pessoas o estão utilizando e com qual objetivo.

O WinBUGS possui um Manual do Usuário, que o acompanha no pacote de instalação, e dois documentos com muitos exemplos. Para o usuário iniciante, recomenda-se como primeira leitura o Tutorial que está inserido no Manual do Usuário. Alguns problemas quanto à programação e mensagens de erro podem ser resolvidos enviando-se um *e-mail* à equipe que idealizou o software ([bugs@mrc-bsu.cam.ac.uk](mailto:bugs@mrc-bsu.cam.ac.uk)). Há também uma lista de discussão de usuários do WinBUGS (<http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A0=bugs>), na qual as pessoas cadastradas podem enviar *e-mails* que serão recebidos por todos os usuários cadastrados.